(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 2 166 099 A2**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
24.03.2010 Bulletin 2010/12

(51) Int Cl.:
C12N 15/82 *(2006.01)*

(21) Application number: **09176168.4**

(22) Date of filing: **07.03.2006**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR

(30) Priority: **08.03.2005 US 659482 P**

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
08168818.6 / 2 045 327
06708664.5 / 1 859 037

(27) Previously filed application:
07.03.2006 EP 08168818

(71) Applicant: **BASF Plant Science GmbH**
67056 Ludwigshafen (DE)

(72) Inventors:
- **Song, Hee-Sook**
  Raleigh, NC 27606 (US)
- **Dammann, Christian**
  Durham, NC 27705 (US)
- **Morra, Marc**
  Bronx, NY 10461 (US)
- **Brown, Jeffrey A.**
  Apex, NC 27502 (US)
- **Xing, Liqun**
  Chapel Hill, NC 27516 (US)
- **Jia, Hongmei**
  Apex, NC 27523 (US)

(74) Representative: **Popp, Andreas**
BASF SE
GVX/B
Carl-Bosch-Str. 38
67056 Ludwigshafen (DE)

Remarks:
This application was filed on 17-11-2009 as a divisional application to the application mentioned under INID code 62.

(54) Expression enhancing intron sequences

(57) The invention relates to methods for the identification and use of introns with gene expression enhancing properties. The teaching of this invention enables the identification of introns causing intron-mediated enhancement (IME) of gene expression. The invention furthermore relates to recombinant expression construct and vectors comprising said IME-introns operably linked with a promoter sequence and a nucleic acid sequence. The present invention also relates to transgenic plants and plant cells transformed with these recombinant expression constructs or vectors, to cultures, parts or propagation material derived there from, and to the use of same for the preparation of foodstuffs, animal feeds, seed, pharmaceuticals or fine chemicals, to improve plant biomass, yield, or provide desirable phenotypes.

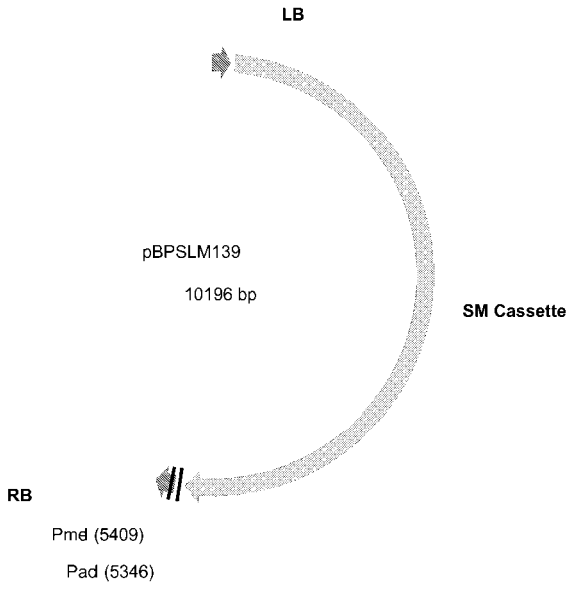

Fig. 1

EP 2 166 099 A2

Description

FIELD OF THE INVENTION

[0001] The invention relates to methods for the identification and use of introns with gene expression enhancing properties. The teaching of this invention enables the identification of introns causing intron-mediated enhancement (IME) of gene expression. The invention furthermore relates to recombinant expression construct and vectors comprising said IME-introns operably linked with a promoter sequence and a nucleic acid sequence. The present invention also relates to transgenic plants and plant cells transformed with these recombinant expression constructs or vectors, to cultures, parts or propagation material derived there from, and to the use of same for the preparation of foodstuffs, animal feeds, seed, pharmaceuticals or fine chemicals, to improve plant biomass, yield, or provide desirable phenotypes.

BACKGROUND OF THE INVENTION

[0002] The aim of plant biotechnology is the generation of plants with advantageous novel properties, such as pest and disease resistance, resistance to environmental stress (*e.g.,* drought), improved qualities (*e.g.,* high yield), or for the production of certain chemicals or pharmaceuticals. Appropriate gene expression rates play an important role in order to obtain the desired phenotypes. The gene expression rate is mainly modulated by the promoter, additional DNA sequence located in the 5' untranscribed and 5' untranslated region and the terminator sequences of a given gene. Promoters are the portion of DNA sequences located at the 5' end a gene which contains signals for RNA polymerases to begin transcription so that a protein synthesis can then proceed. Regulatory DNA sequences positioned in the 5' untranscribed region modulate gene expression in response to specific biotic (e.g. pathogen infection) or abiotic (e.g. salt-, heat-, drought-stress) stimuli. Furthermore, other so called "enhancer" sequences have been identified that elevate the expression level of nearby located genes in a position and orientation independent manner.

[0003] Beside the elements located on the untranscribed regions of a gene (e.g. promoter, enhancer), it is documented in a broad range of organisms (e.g. nematodes, insects, mammals and plants) that some introns have gene expression enhancing properties. In plants, the inclusion of some introns in gene constructs leads to increased mRNA and protein accumulation relative to constructs lacking the intron. This effect has been termed "intron mediated enhancement" (IME) of gene expression (Mascarenhas et al., (1990) Plant Mol. Biol. 15:913-920). Introns known to stimulate expression in plants have been identified in maize genes (*e.g. tubA1, Adh1, Sh1, Ubi1* (Jeon et al. (2000) Plant Physiol. 123:1005-1014; Callis et al. (1987) Genes Dev. 1:1183-1200; Vasil et al. (1989) Plant Physiol 91:1575-1579; Christiansen et al. (1992) Plant Mol. Biol. 18:675-689]) and in rice genes (*e.g. salT, tpi* [McElroy et al. (1990) Plant Cell 2: 163-171; Xu et al. (1994) Plant Physiol 106:459-467]). Similarly, introns from dicotyledonous plant genes like those from petunia (*e.g. rbcS*), potato (*e.g. st-ls1*) and from *Arabidopsis thaliana* (*e.g. ubq3* and *pat1*) have been found to elevate gene expression rates (Dean et al. (1989) Plant Cell 1:201-208; Leon et al. (1991) Plant Phyisiol. 95:968-972; Norris et al. (1993) Plant Mol Biol 21: 895-906; Rose and Last (1997) Plant J 11:455-464). It has been shown that deletions or mutations within the splice sites of an intron reduce gene expression, indicating that splicing might be needed for IME (Mascarenhas et al. (1990) Plant Mol Biol 15:913-920; Clancy and Hannah (2002) Plant Physiol 130:918-929). However, that splicing per se is not required for a certain IME in dicotyledonous plants has been shown by point mutations within the splice sites of the *pat1* gene from *A.thaliana* (Rose and Beliakoff (2000) Plant Physiol 122:535-542).

[0004] Enhancement of gene expression by introns is not a general phenomenon because some intron insertions into recombinant expression cassettes fail to enhance expression (e.g. introns from dicot genes (rbcS gene from pea, phaseolin gene from bean and the *stls*-1 gene from *Solanum tuberosum*) and introns from maize genes (*adh1* gene the ninth intron, hsp81 gene the first intron)) (Chee et al. (1986) Gene 41:47-57; Kuhlemeier et al. (1988) Mol Gen Genet 212:405-411; Mascarenhas et al. (1990) Plant Mol Biol 15:913-920; Sinibaldi and Mettler (1992) In WE Cohn, K Moldave, eds, Progress in Nucleic Acid Research and Molecular Biology, Vol 42. Academic Press, New York, pp 229-257; Vancanneyt et al. 1990 Mol Gen Gent 220:245-250). Therefore, not each intron can be employed in order to manipulate the gene expression level of alien genes or endogenous genes in transgenic plants. What characteristics or specific sequence features must be present in an intron sequence in order to enhance the expression rate of a given gene is not known in the prior art and therefore from the prior art it is not possible to predict whether a given plant intron, when used heterologously, will cause IME.

[0005] The introduction of a foreign gene into a new plant host does not always result in a high expression of the incoming gene. Furthermore, if dealing with complex traits, it is sometimes necessary to modulate several genes with spatially or temporarily different expression pattern. Introns can principally provide such modulation. However multiple use of the same intron in one plant has shown to exhibit disadvantages. In those cases it is necessary to have a collection of basic control elements for the construction of appropriate recombinant DNA elements. However, the available collection of introns with expression enhancing properties is limited and alternatives are needed.

[0006] Thus, there is still a growing demand for basic control elements including promoters, regulatory sequences

(*e.g.,* inducible elements, enhancers) or intron sequences that have an impact on gene expression rates. It is therefore an objective of the present invention, to provide a highly reproducible and reliable method for the identification of introns with expression enhancing properties.

[0007] This objective is achieved by the methods provided within this invention.

SUMMARY OF THE INVENTION

[0008] A first subject matter of the invention therefore relates to a method for identifying an intron with expression enhancing properties in plants comprising selecting an intron from a plant genome, wherein said intron is characterized by at least the following features

I) an intron length shorter than 1,000 base pairs, and

II) presence of a 5' splice site comprising the dinucleotide sequence 5'-GT-3' (SEQ ID NO: 78), and

III) presence of a 3' splice site comprising the trinucleotide sequence 5'-CAG-3' (SEQ ID NO: 79), and

IV) presence of a branch point resembling the consensus sequence 5'-CURAY-3' (SEQ ID NO:75) upstream of the 3'splice site, and

V) an adenine plus thymine content of at least 40% over 100 nucleotides downstream from the 5' splice site, and

VI) an adenine plus thymine content of at least 50% over 100 nucleotides upstream from the 3' splice site, and

VII) an adenine plus thymine content of at least 50%, and a thymine content of at least 30% over the entire intron.

[0009] In another embodiment, the invention relates to a method for enriching the number of introns with expression enhancing properties in plants in a population of plant introns to a percentage of at least 50% of said population, said method comprising selecting introns from said population, wherein said introns are characterized by at least the following features

I) an intron length shorter than 1,000 base pairs, and

II) presence of a 5' splice site comprising the dinucleotide sequence 5'-GT-3' (SEQ ID NO: 78), and

III) presence of a 3' splice site comprising the trinucleotide sequence 5'-CAG-3' (SEQ ID NO: 79), and

IV) presence of a branch point resembling the consensus sequence 5'-CURAY-3' (SEQ ID NO:75) upstream of the 3'splice site, and

V) an adenine plus thymine content of at least 40% over 100 nucleotides downstream from the 5' splice site, and

VI) an adenine plus thymine content of at least 50% over 100 nucleotides upstream from the 3' splice site, and

VII) an adenine plus thymine content of at least 50%, and a thymine content of at least 30% over the entire intron.

[0010] Preferably, the population of plant introns chosen for the enrichment of introns with gene expression enhancing properties in plants comprises substantially all introns of a plant genome represented in a genomic DNA sequence database or a plant genomic DNA library.

[0011] In a preferred embodiment, the intron with gene expression enhancing properties in plants ("IME-intron") is selected by the method of the invention for identifying IME-introns or the method of the invention for enriching the number of IME-introns in a population of plant introns. Preferably, said intron is selected from the group consisting of introns located between two protein encoding exons or introns located within the 5' untranslated region of the corresponding gene.

[0012] In a particularly preferred embodiment, the IME-intron is identified or enriched by one of the inventive methods from a group or population of genes representing the 10% fraction of genes with the highest expression rate in a gene expression analysis experiment performed using a plant cell, plant tissue or a whole plant.

[0013] The invention furthermore relates to a method wherein the gene sequence information used for the identification or enrichment of IME-introns is present in a DNA sequence database and the selection steps for identifying or enriching

said introns are performed using an automated process, preferably by using a computer device and an algorithm that defines the instructions needed for accomplishing the selection steps for identifying or enriching said introns.

[0014] Additionally, the invention relates to computer algorithm that defines the instructions needed for accomplishing the selection steps for identifying or enriching IME-introns from a plant genome or a population of introns selected from the group consisting of introns located between two protein encoding exons, and/or introns located within the 5' untranslated region of the corresponding gene and/or introns located in the DNA sequences of genes representing the 10% fraction of genes with the highest expression rate in a gene expression analysis experiment performed using a plant cell, plant tissue and/or a whole plant.

[0015] The invention also relates to the computer device or data storage device comprising an algorithm as described above.

[0016] In a preferred embodiment, the invention relates to methods for isolating, providing or producing IME-introns comprising the steps of performing an identification or enrichment of IME-introns as described above and providing the sequence information of said IME-introns identified or enriched, and providing the physical nucleotide sequence of said identified or enriched introns and evaluating the gene expression enhancing properties of the isolated introns in an *in vivo* or *in vitro* expression experiment, and isolating the IME-introns from the population of introns tested in the *in vivo* or *in vitro* expression experiment. Preferably, the evaluation of the gene expression enhancing properties of the IME-intron is done in a plant cell and wherein IME-intron enhances the expression of a given nucleic acid at least twofold.

[0017] An additional subject matter of the invention relates to a recombinant DNA expression construct comprising at least one promoter sequence functioning in plants cells, at least one nucleic acid sequence and at least one intron selected from the group consisting of the sequences described by SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22, and functional equivalents thereof, wherein said promoter sequence and at least one of said intron sequences are functionally linked to said nucleic acid sequence and wherein said intron is heterologous to said nucleic acid sequence or to said promoter sequence.

[0018] Furthermore, the invention relates to recombinant expression constructs comprising at least one promoter sequence functioning in plants cells, at least one nucleic acid sequence and at least one functional equivalents of an intron described by any of sequences SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22, wherein said functional equivalent comprises the functional elements of an intron and is characterized by

a) a sequence having at least 50 consecutive base pairs of the intron sequence described by any of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22, or

b) having an identity of at least 80% over a sequence of at least 95 consecutive nucleic acid base pairs to a sequences described by any of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22, or

c) hybridizing under high stringent conditions with a nucleic acid fragment of at least 50 consecutive base pairs of a nucleic acid molecule described by any of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22,

wherein said promoter sequence and at least one of said intron sequences are functionally linked to said nucleic acid sequence and wherein said intron is heterologous to said nucleic acid sequence or to said promoter sequence.

[0019] In another embodiment, the recombinant DNA expression construct of the invention further contains one or more additional regulatory sequences functionally linked to promoter. Those regulatory sequences can be selected from the group consisting of heat shock responsive-, anaerobic responsive-, pathogen responsive-, drought responsive-, low temperature responsive-, ABA responsive-elements, 5' untranslated gene region, 3' untranslated gene region, transcription terminators, polyadenylation signals and enhancers.

[0020] The nucleic acid sequence of the inventive recombinant DNA expression construct may result in the expression of a protein and/or sense, antisense or double-stranded RNA encoded by said nucleic acid sequence.

[0021] In another embodiment, the nucleotide sequence encoding the transgenic expression construct of the invention is double-stranded. In yet another embodiment, the nucleotide sequence encoding the transgenic expression construct of the invention is single-stranded.

[0022] In yet another alternative embodiment of the invention, the recombinant expression construct comprises a nucleic acid sequence encoding for a selectable marker protein, a screenable marker protein, a anabolic active protein, a catabolic active protein, a biotic or abiotic stress resistance protein, a male sterility protein or a protein affecting plant agronomic characteristics.

[0023] The invention relates furthermore to vectors containing a transgenic expression construct of the invention. Additionally, the invention relates to transgenic cells or transgenic non-human-organisms like bacteria, fungi, yeasts or plants comprising an expression vector containing a transgenic expression construct of the invention. In a preferred embodiment, the transgenic cell or transgenic non-human organism transformed with an expression construct of the

invention is a monocotyledonous plant or is derived from such a plant. In a yet more preferred embodiment, the monocotyledonous plant is selected from the group consisting of the genera *Hordeum, Avena, Secale, Triticum, Sorghum, Zea, Saccharum,* and *Oryza.* Further embodiments of the invention relate to cell cultures, parts or propagation material derived from non-human-organisms like bacteria, fungi, yeasts and/or plants, preferably monocotyledonous plants, most preferably plants selected from the group consisting of the genera *Hordeum, Avena, Secale, Triticum, Sorghum, Zea, Saccharum,* and *Oryza,* transformed with the inventive vectors or containing the inventive recombinant expression constructs.

[0024] The invention furthermore relates to a method for providing an expression cassette for enhanced expression of a nucleic acid sequence in a plant or a plant cell, comprising the step of functionally linking at least one sequence selected from the group consisting of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 to said nucleic acid sequence.

[0025] The invention further relates to a method for enhancing the expression of a nucleic acid sequence in a plant or a plant cell, comprising functionally linking at least one sequence selected from the group consisting of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 to said nucleic acid sequence.

[0026] An additional embodiment of the invention relates to a method

a) for providing an expression cassette for enhanced expression of a nucleic acid sequence in a plant or a plant cell, or
b) for enhancing the expression of a nucleic acid sequence in a plant or a plant cell said method comprising functionally linking at least one sequence selected from the group consisting of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 to said nucleic acid sequence, wherein furthermore a promoter sequence functional in plants is linked to said nucleic acid sequence.

[0027] Preferably, at least one sequence selected from the group consisting of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 is linked to a nucleic acid sequence by insertion into the plant genome via homologous recombination. Preferably, said homologous recombination is comprising at least the steps of

a) providing *in vivo* or *in vitro* a DNA construct comprising said intron flanked by sequences ("recombination substrate") allowing homologous recombination into a pre-existing expression cassette between the promoter and the nucleic acid of said expression cassette, and
b) transforming a recipient plant cell comprising said cassette of step a) and regenerating a transgenic plant, wherein said intron has been inserted into the genome of said plant. Preferably, the site of integration into the genome of said plant is determined by the DNA sequence of the recombination substrate of step a), wherein said sequence sharing sufficient homology (as defined herein) with said genomic target DNA sequence allowing the sequence specific integration via homologous recombination at said genomic target DNA locus.

[0028] In a preferred embodiment of the invention, said recipient plant or plant cell is a monocotyledonous plant or plant cell, more preferably a plant or plant cell selected from the group consisting of the genera *Hordeum, Avena, Secale, Triticum, Sorghum, Zea, Saccharum,* and *Oryza,* most preferably a maize plant.

[0029] Preferably, the nucleic acid sequence to which one of the inventive intron is functionally linked, encodes for a selectable marker protein, a screenable marker protein, an anabolic active protein, a catabolic active protein, a biotic or abiotic stress resistance protein, a male sterility protein or a protein affecting plant agronomic characteristics and/or a sense, antisense, or double-stranded RNA.

[0030] Additionally, the invention relate to the use of a transgenic organism of the invention or of cell cultures, parts of transgenic propagation material derived there from for the production of foodstuffs, animal feeds, seeds, pharmaceuticals or fine chemicals.

[0031] The invention furthermore relates to a recombinant DNA expression construct comprising

a) at least one promoter sequence functioning in plants or plant cells, and

b) at least one intron selected from the group of introns with expression enhancing properties in plants or plant cells characterized by at least the following features

I) an intron length shorter than 1,000 base pairs, and

II) presence of a 5' splice site comprising the dinucleotide sequence 5'-GT-3' (SEQ ID NO: 78), and

III) presence of a 3' splice site comprising the trinucleotide sequence 5'-CAG-3' (SEQ ID NO: 79), and

IV) presence of a branch point resembling the consensus sequence 5'-CURAY-3' (SEQ ID NO: 75) upstream of the 3'splice site, and

V) an adenine plus thymine content of at least 40% over 100 nucleotides downstream from the 5' splice site, and

VI) an adenine plus thymine content of at least 50% over 100 nucleotides upstream from the 3' splice site, and

VII) an adenine plus thymine content of at least 55%, and a thymine content of at least 30% over the entire intron, and

c) at least one nucleic acid sequence, wherein said promoter sequence and at least one of said intron sequences are functionally linked to said nucleic acid sequence and wherein said intron is heterologous to said nucleic acid sequence and/or to said promoter sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

[0032]

Fig. 1 Map of pBPSMM291 (SEQ ID NO: 109) This vector comprises the maize ubiquitin promoter, followed by the BPSI.1, then the GUSint ORF (including the potato invertase [PIV]2 intron to prevent bacterial expression), followed by nopaline synthase (NOS) terminator. This vector contains the attL1 and attL2 sites to make it compatible with modification via the Gateway® cloning Technology from Invitrogen™. This vector is based on the pUC based expression vector pBPSMM267. The *Xma*I-*Rsr*II digested BPSI.1 PCR product was ligated into the *Xma*I-*Rsr*II digested pBPSMM267 to create pBPSMM291. The vectors pBPSMM293, pBPSMM294 and pBPSMM295 have been created accordingly (see table 6 and 1.6.1).

Fig. 2 Map of pBPSMM305 (SEQ ID NO:110) The expression vector pBPSMM305 comprises the maize lactate dehydro- genase (LDH) promoter without intron driving expression of the GUSint ORF (including the potato invertase [PIV]2 intron to prevent bacterial expression), fol- lowed by the NOS terminator. This vector has been used to create the pUC based expression vectors pBPSJB041, pBPSJB042, pBPSJB043, pBPSJB044, pBPSJB045, pBPSJB046 and pBPSJB050 (see examples 2.3).

Fig. 3 Map of pBPSMM350 (SEQ ID NO:111): The vector pBPSMM350 comprises the maize ubiquitin promoter, followed by the BPSI.1, then the GUSint ORF (including the potato invertase [PIV]2 intron to prevent bacterial expression), followed by nopaline synthase (NOS) terminator. The expression cassette has been transferred from the vector pBPSMM291 us- ing the Gateway® cloning Technology from Invitrogen™. The vectors pBPSMM353, pBPSMM312 and pBPSMM310 have been created accordingly (see table 6 and example 1.6.2).

Fig. 4 Map of pBPSLM139 (SEQ ID NO:112): The vector pBPSLM139 comprises the selectable marker expression cassette. In order to produce the vectors pBPSLI017 to pBPSLI023, *Pme*I/*Pac*I fragments have been isolated from the vectors pBPSJB-042, -043, -044, -045, 046 and 050 and cloned into the *Pme*I-*Pac*I digested pBPSLM130 (see example 2.3 and 2.4)

Fig. 5a-f: Computer algorithm for retrieving sequence information from NCBI genebank file.

Fig. 6 Transgenic plants containing promoter constructs with BPSI.1 intron (all but pBPSLM229) or BPSI.5 intron (only pBPSLM229) were tested for GUS expression at 5-leaf (A), flowering (B) and seed set (C) stages. Shown are examples of typical staining patterns obtained from at least 15 independent events. All samples were stained for 16 hours in GUS solution. Promoters in the constructs are: rice chloroplast protein 12 (Os.CP12; pBPSMM355), the maize hydroxyproline-rich glycoprotein (Zm.HRGP; pBPSMM370), the rice p-caffeoyl- CoA 3-O-methyltransferase (Os.CCoAMT1; pBPSMM358), the maize Globulin- 1 promoter W64A (Zm.Glb1; EXS1025), the putative Rice H+-transporting ATP synthase promoter (Os.V-ATPase; pBPSMM369), Zm.LDH (pBPSMM357), the rice C-8,7 sterol isomerase promoter (Os.C8,7 SI; pBPSMM366), the rice Late Embryogenesis Abundant Protein promoter (Os.Lea; pBPSMM371), and the maize lactate dehydrogenase promoter (ZM.LDH; pBPSLM229)..

GENERAL DEFINITIONS

[0033] It is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, plant species or genera, constructs, and reagents described as such It must be noted that as used herein and in the appended claims, the singular forms "a" and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a vector" is a reference to one or more vectors and includes equivalents thereof known to those skilled in the art.

[0034] About: the term "about" is used herein to mean approximately, roughly, around, or in the region of. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 20 percent, preferably 10 percent up or down (higher or lower). As used herein, the word "or" means any one member of a particular list.

[0035] *Agrobacterium:* refers to a soil-borne, Gram-negative, rod-shaped phytopathogenic bacterium which causes crown gall. The term *"Agrobacterium"* includes, but is not limited to, the strains *Agrobacterium tumefaciens,* (which typically causes crown gall in infected plants), and *Agrobacterium rhizogenes* (which causes hairy root disease in infected host plants). Infection of a plant cell with *Agrobacterium* generally results in the production of opines (*e.g.,* nopaline, agropine, octopine etc.) by the infected cell. Thus, *Agrobacterium* strains which cause production of nopaline (*e.g.,* strain LBA4301, C58, A208) are referred to as "nopaline-type" *Agrobacteria*; *Agrobacterium* strains which cause production of octopine (*e.g.,* strain LBA4404, Ach5, B6) are referred to as "octopine-type" *Agrobacteria*; and *Agrobacterium* strains which cause production of agropine (*e.g.,* strain EHA105, EHA101, A281) are referred to as "agropine-type" *Agrobacteria.*

[0036] Algorithm: as used herein refers to the way computers process information, because a computer program is essentially an algorithm that tells the computer what specific steps to perform (in what specific order) in order to carry out a specified task, such as identification of coding regions of a set of genes. Thus, an algorithm can be considered to be any sequence of operations that can be performed by a computer system. Typically, when an algorithm is associated with processing information, data is read from an input source or device, written to an output sink or device, and/or stored for further use. For any such computational process, the algorithm must be rigorously defined: specified in the way it applies in all possible circumstances that could arise. That is, any conditional steps must be systematically dealt with, case-by-case; the criteria for each case must be clear (and computable). Because an algorithm is a precise list of precise steps, the order of computation will almost always be critical to the functioning of the algorithm. Instructions are usually assumed to be listed explicitly, and are described as starting 'from the top' and going 'down to the bottom', an idea that is described more formally by flow of control. In computer applications, a script is a computer program that automates the sort of task that a user might otherwise do interactively at the keyboard. Languages that are largely used to write such scripts are called scripting languages. Many such languages are quite sophisticated, and have been used to write elaborate programs, which are often still called scripts even if they go well beyond automating simple sequences of user tasks. Computer languages are created for varying purposes and tasks different kinds and styles of programming. Scripting programming languages (commonly called scripting languages or script languages) are computer programming languages designed for "scripting" the operation of a computer. Early script languages were often called batch languages or job control languages.

[0037] Examples for script languages are: ACS, ActionScript, Active Server Pages (ASP), AppleScript, Awk, BeanShell (scripting for Java), bash, Brain, CobolScript, csh, Cold-Fusion, Dylan, Escapade (server side scripting), Euphoria, Groovy, Guile, Haskell, HyperTalk, ICI, IRC script, JavaScript, mIRC script, MS-DOS batch, Nwscript, Perl, PHP, Pike, ScriptBasic.

[0038] Antisense: is understood to mean a nucleic acid having a sequence complementary to a target sequence, for example a messenger RNA (mRNA) As used herein, the terms "complementary" or "complementarity" are used in reference to nucleotide sequences related by the base-pairing rules. For example, the sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

[0039] Sense: is understood to mean a nucleic acid having a sequence that is homologous or identical to a target sequence, for example a sequence which is bound by a protein factor of the spliceosome.

[0040] Bombarding, "bombardment and "biolistic bombardment": refer to the process of accelerating particles (microprojectiles) towards a target biological sample (*e.g.,* cell, tissue, etc.) to effect wounding of the cell membrane of a cell in the target biological sample and/or entry of the particles into the target biological sample. Methods for biolistic bombardment are known in the art (*e.g.,* US 5,584,807, the contents of which are herein incorporated by reference), and are commercially available (*e.g.,* the helium gas-driven microprojectile accelerator (PDS-1000/He) (BioRad).

[0041] Cell: refers to a single cell. The term "cells" refers to a population of cells. The population may be a pure

population comprising one cell type. Likewise, the population may comprise more than one cell type. In the present invention, there is no limit on the number of cell types that a cell population may comprise. The cells may be synchronize or not synchronized, preferably the cells are synchronized.

[0042] Chromosomal DNA or chromosomal DNA-sequence: is to be understood as the genomic DNA of the cellular nucleus independent from the cell cycle status. Chromosomal DNA might therefore be organized in chromosomes or chromatids, they might be condensed or uncoiled. An insertion into the chromosomal DNA can be demonstrated and analyzed by various methods known in the art like *e.g.,* polymerase chain reaction (PCR) analysis, Southern blot analysis, fluorescence *in situ* hybridization (FISH), and *in situ* PCR.

[0043] Coding region or coding sequence (CDS): when used in reference to a gene refers to the nucleotide sequences which encode the amino acids found in the nascent polypeptide as a result of translation of a mRNA molecule. The coding region is bounded, in eucaryotes, on the 5'-side by the nucleotide triplet "ATG" which encodes the initiator methionine and on the 3'-side by one of the three triplets, which specify stop codons (*i.e.,* TAA, TAG, TGA).

[0044] Complement of a nucleic acid sequence: as used herein refers to a nucleotide sequence whose nucleic acids show total complementarity to the nucleic acids of the nucleic acid sequence.

[0045] Decile: when used in connection with statistical data is any of the 10 values that divide sorted data into 10 equal parts, so that each part represents 1/10th of the sample or population. Thus, the 1 st decile cuts off lowest 10% of data, the 9th decile cuts off lowest 90% or the highest 10% of data. A quartile is any of the three values which divide the sorted data set into four equal parts, so that each part represents 1/4th of the sample or population (third quartile = upper quartile = cuts off highest 25% of data, or lowest 75% = *75th* percentile). A percentile is any of the 99 values that divide the sorted data into 100 equal parts, so that each part represents 1/100th of the sample or population. Thus, the 1st percentile cuts off lowest 1% of data, the 98th percentile cuts off lowest 98% of data and the 25^{th} percentile cuts off lowest 25% of data.

[0046] DNA databases: in the field of bioinformatics, a DNA sequence database is a large collection of DNA sequences stored on a computer. A database can include sequences from only one organism, or it can include sequences from all organisms whose DNA has been sequenced.

[0047] Enrichment or enriching: when used in connection with the selection of inventive introns refers to an increase in the success rate of identifying introns with gene expression enhancing properties within a population of introns (e.g. a population of introns representing all introns of a plant genome present in a genomic DNA sequence database). The enrichment is achieved by reducing the number of candidate introns by using the inventive method and the inventive selection criteria. If, as an example, the success rate of identifying an intron with expression enhancing properties from a given population of introns - by using the herein described methods for measuring gene expression enhancement- is one out of ten analyzed introns, enrichment has to be understood as an increase in the number of identified introns with gene expression enhancing properties -by using the inventive method- to at least five out of ten analyzed introns. Therefore, the number of introns needed to be analyzed in order to identify one inventive intron is reduced to two introns by using the inventive method as a preselection or filtering process.

[0048] Evaluation of the expression enhancing properties: of an intron can be done using methods known in the art. For example, a candidate intron sequence whose gene expression enhancing effect is to be determined can be inserted into the 5'UTR of a nucleic acid sequence encoding for a reporter gene (*e.g.,* a visible marker protein, a selectable marker protein) under control of an appropriate promoter active in plants or plant cells to generate a reporter vector. The reporter vector and an identical control reporter vector lacking the candidate intron can be introduced into a plant tissue using methods described herein, and the expression level of the reporter gene, in dependence of the presence of the candidate intron, can be measured and compared (*e.g.,* detecting the presence of encoded mRNA or encoded protein, or the activity of a protein encoded by the reporter gene). An intron with expression enhancing properties will result in a higher expression rate than a reference value obtained with an identical control reporter vector lacking the candidate intron under otherwise unchanged conditions.

[0049] The reporter gene may express visible markers. Reporter gene systems which express visible markers include β-glucuronidase and its substrate (X-Gluc), luciferase and its substrate (luciferin), and β-galactosidase and its substrate (X-Gal) which are widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes (1995) Methods Mol Biol 55:121-131). The assay with β-glucuronidase (GUS) being very especially preferred (Jefferson et al., GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. (1987) Dec 20;6(13):3901-3907). β-glucuronidase (GUS) expression is detected by a blue color on incubation of the tissue with 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid. The selectable marker gene may confer antibiotic or herbicide resistance. Examples of reporter genes include, but are not limited to, the *dhfr* gene, which confers resistance to methotrexate (Wigler (1980) Proc Natl Acad Sci 77: 3567-3570); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin (1981) J. Mol. Biol. 150:1-14) and *als* or *pat,* which confer resistance to chlorsulfuron and phosphinotricin acetyl transferase, respectively.

[0050] Expect value when used in the context of DNA sequence alignments or DNA sequence database searches

refers to the number of times a certain match or a better one would be expected to occur purely by chance in a search of the entire database. Thus, the lower the Expect value, the greater the similarity between the input sequence and the match. The Expect value (E) is a parameter that describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. It decreases exponentially with the Similarity Score (S) that is assigned to a match between two sequences. The higher the score, the lower the E value. Essentially, the E value describes the random background noise that exists for matches between sequences. The Expect value is used as a convenient way to create a significance threshold for reporting results. An E value of 1 assigned to a hit can be interpreted as meaning that in a database of the current size you might expect to see 1 match with a similar score simply by chance. The E-value is influenced by: a) length of sequence (the longer the query the lower the probability that it will find a sequence in the database by chance), b) size of database (the larger the database the higher the probability that the query will find a match by chance), c) the scoring matrix (the less stringent the scoring matrix the higher the probability that the query will find a sequence in the database by chance).

[0051] Expressed sequence tag (EST): refers to a cDNA sequence that has been obtained from a single pass terminal DNA sequencing. An EST sequence denotes a sequence that is derived from a transcript, and hence from a gene that is transcribed.

[0052] Expressible nucleic acid sequence: as used in the context of this invention is any nucleic acid sequence that is capable of being transcribed into RNA (e.g. mRNA, antisense RNA, double strand forming RNA etc.) or translated into a particular protein.

[0053] Expression: refers to the biosynthesis of a gene product. For example, in the case of a structural gene, expression involves transcription of the structural gene into mRNA and - optionally - the subsequent translation of mRNA into one or more polypeptides.

[0054] Functional equivalents: with regard to the inventive introns has to be understood as natural or artificial mutations of said introns described in any of the SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22. Mutations can be insertions, deletions or substitutions of one or more nucleic acids that do not diminish the expression enhancing properties of said introns. These functional equivalents having a identity of at least 80%, preferably 85%, more preferably 90%, most preferably more than 95%, very especially preferably at least 98% identity - but less then 100% identity to the intron sequences as described by any of the SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22, wherein said identity is determined over a sequence of at least 95 consecutive base pairs, preferably at least 150 consecutive base pairs, more preferably at least 200 consecutive base pairs of the sequence as described by any of the SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 and having essentially the same IME effect characteristics as the intron sequences as shown in any of the SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22.

[0055] Functional equivalents are in particular homologs of said introns derived from other plant species. Homologs when used in reference to introns refers to introns with expression enhancing properties isolated from a genomic nucleic acid sequence that encodes for a protein

(i) sharing more than 60%, preferably 65%, 70%, 75%, 80%, more preferably 85%, 90%, 95% or most preferably more than 95% sequence identity on amino acid level with proteins that are encoded by genes from which the inventive introns with the SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 have been isolated, or

(ii) catalyzing the same enzymatic reaction as the proteins encoded by genes from which the inventive introns SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 have been isolated, or

(iii) showing comparable spatial and temporal expression pattern as the proteins encoded by genes from which the inventive introns SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 have been isolated.

[0056] "Functional equivalents" as described above might have, compared with the inventive introns a reduced or increased gene expression enhancing effect. In this context, the gene expression enhancing effect of the functional equivalent intron is at least 50% higher, preferably at least 100% higher, especially preferably at least 300% higher, very especially preferably at least 500% higher than a reference value obtained with any of the introns shown in SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 under otherwise unchanged conditions.

[0057] Functionally linked or operably linked: is to be understood as meaning, for example, the sequential arrangement of a regulatory element (*e.g.* a promoter) with a nucleic acid sequence to be expressed and, if appropriate, further regulatory elements (such as *e.g.,* a terminator) in such a way that each of the regulatory elements can fulfill its intended function to allow, modify, facilitate or otherwise influence expression of said nucleic acid sequence. The expression may result depending on the arrangement of the nucleic acid sequences in relation to sense or antisense RNA. To this end, direct linkage in the chemical sense is not necessarily required. Genetic control sequences such as, for example, enhancer sequences, can also exert their function on the target sequence from positions that are further away, or indeed from other DNA molecules. The terms "functionally linked", "operably linked," "in operable combination," and "in operable

order" as used herein with reference to an inventive intron with gene expression enhancing properties refers to the linkage of at least one of said introns to a nucleic acid sequences in a way that the expression enhancing effect is realized and, if functional splice sites have been included, that the intron can be spliced out by the cell factors responsible for the splicing procedure. In a preferred embodiment of the present invention, the intron is introduced into the 5' non coding region of a nucleic acid sequence. Inventive expression constructs, wherein an inventive intron is functionally linked to an nucleic acid sequence are shown in the examples. More preferred arrangements are those in which an intron functioning in intron mediated expression enhancement is inserted between a promoter and a nucleic acid sequence, preferably into the transcribed nucleic acid sequence, or in case of a nucleic acid sequence encoding for a protein, into the 5' untranslated region of a nucleic acid sequence. The distance between the promoter sequence and the nucleic acid sequence to be expressed recombinantly is preferably less than 200 base pairs, especially preferably less than 100 base pairs, very especially preferably less than 50 base pairs. Operable linkage, and an expression cassette, can be generated by means of customary recombination and cloning techniques as are described, for example, in Maniatis T, Fritsch EF and Sambrook J (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY), in Silhavy TJ, Berman ML and Enquist LW (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY), in Ausubel FM et al. (1987) Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience and in Gelvin et al. (1990) In: Plant Molecular Biology Manual. However, further sequences which, for example, act as a linker with specific cleavage sites for restriction enzymes, or as a signal peptide, may also be positioned between the two sequences. The insertion of sequences may also lead to the expression of fusion proteins. Preferably, the expression construct, consisting of a linkage of promoter, intron and nucleic acid sequence to be expressed, can exist in a vector-integrated form and be inserted into a plant genome, for example by transformation.

[0058] Gene: refers to a coding region operably linked to appropriate regulatory sequences capable of regulating the expression of the polypeptide in some manner. A gene includes untranslated regulatory regions of DNA (*e.g.,* promoters, enhancers, repressors, etc.) preceding (upstream) and following (downstream) the coding region (open reading frame, ORF) as well as, where applicable, intervening sequences (*i.e.,* introns) between individual coding regions (*i.e.,* exons). Genes may also include sequences located on both the 5'- and 3'-end of the sequences, which are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5'-flanking region may contain regulatory sequences such as promoters and enhancers, which control or influence the transcription of the gene. The 3'-flanking region may contain sequences, which direct the termination of transcription, posttranscriptional cleavage and polyadenylation.

[0059] Gene expression enhancing properties, gene expression enhancing effect or intron mediated gene expression enhancement (IME): when made in reference to an intron sequence refers to the ability of the intron to enhance quantitatively the expression level of a nucleic acid sequence (e.g. a gene) that is part of an recombinant/transgenic DNA expression cassette (as defined herein), measured on the basis of the transcribed RNA, mRNA, protein amount or protein activity compared to the otherwise identical expression construct lacking the intron under otherwise unchanged conditions. Gene expression enhancing properties in plants: refers to an intron that is able to enhance quantitatively the expression level of a plant derived nucleic acid sequence in a plant or plant cell and the enhancement of gene expression rate of a non-plant derived nucleic acid in a plant or a plant cell compared to the otherwise identical expression construct lacking the intron under otherwise unchanged conditions. In a preferred embodiment of the invention, the expression enhancing effect is understood as an increase in the RNA steady state level, the protein steady state level or the protein activity of a nucleic acid sequence or the corresponding protein (e.g. a reporter gene or protein) of at least 50%, or at least 100%, or at least 200%, 300%, 400% or at least 500%, 600%, 700%, 800%, 900% or at least 1,000%, or more than 1,000% compared to the otherwise identical expression construct lacking the intron under otherwise unchanged conditions. Furthermore expression enhancing effect or intron mediated enhancement has to be understood as the ability of an intron to change the tissue, organ or cell specific expression pattern of a nucleic acid sequence (e.g. a gene) that is part of an inventive expression cassette. Changing the tissue, organ or cell specific expression pattern of a nucleic acid sequence that is part of an inventive expression cassette refers to the fact that due to the presence of an inventive intron, the expression level (mRNA or encoded protein steady state level, or the activity of a protein) of the respective gene is increased above the detection threshold of the used detection method.

[0060] Gene silencing: can be realized by antisense or double-stranded RNA or by co-suppression (sense-suppression). The skilled worker knows that he can use alternative cDNA or the corresponding gene as starting template for suitable antisense constructs. The "antisense" nucleic acid is preferably complementary to the coding region of the target protein or part thereof. However, the "antisense" nucleic acid may also be complementary to the non-coding region or part thereof. Starting from the sequence information on a target protein, an antisense nucleic acid can be designed in the manner with which the skilled worker is familiar, taking into consideration Watson's and Crick's rules of base pairing. An antisense nucleic acid can be complementary to the entire or part of the nucleic acid sequence of a target protein.

[0061] Likewise encompassed is the use of the above-described sequences in sense orientation, which, as is known to the skilled worker, can lead to co-suppression (sense-suppression). It has been demonstrated that expression of

sense nucleic acid sequences can reduce or switch off expression of the corresponding gene, analogously to what has been described for antisense approaches (Goring (1991) Proc. Natl Acad. Sci. USA 88:1770-1774; Smith (1990) Mol. Gen. Genet. 224:447-481; Napoli (1990) Plant Cell 2:279-289;Van der Krol (1990) Plant Cell 2:291-299). In this context, the construct introduced may represent the gene to be reduced fully or only in part. The possibility of translation is not necessary. Especially preferred is the use of gene regulation methods by means of double-stranded RNAi ("double-stranded RNA interference"). Such methods are known to the person skilled in the art (*e.g.,* Matzke 2000; Fire 1998; WO 99/32619; WO 99/53050; WO 00/68374; WO 00/44914; WO 00/44895; WO 00/49035; WO 00/63364). The processes and methods described in the references stated are expressly referred to.

[0062] Genome and genomic DNA of an organism as used herein is the whole hereditary information of an organism that is encoded in the DNA (or, for some viruses, RNA). This includes both the genes and the non-coding sequences. Said genomic DNA comprises the DNA of the nucleus (also referred to as chromosomal DNA) but also the DNA of the plastids (*e.g.,* chloroplasts) and other cellular organelles (*e.g.,* mitochondria). Preferably the terms genome or genomic DNA is referring to the chromosomal DNA of the nucleus. The term "chromosomal DNA" or "chromosomal DNA-sequence" is to be understood as the genomic DNA of the cellular nucleus independent from the cell cycle status. Chromosomal DNA might therefore be organized in chromosomes or chromatids, they might be condensed or uncoiled. An insertion into the chromosomal DNA can be demonstrated and analyzed by various methods known in the art like *e.g.,* polymerase chain reaction (PCR) analysis, Southern blot analysis, fluorescence *in situ* hybridization (FISH), and *in situ* PCR.

[0063] Heterologous: with respect to a nucleic acid sequence refers to a nucleotide sequence, which is ligated to a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature.

[0064] Hybridizing: as used herein includes "any process by which a strand of nucleic acid joins with a complementary strand through base pairing." (Coombs 1994, Dictionary of Biotechnology, Stockton Press, New York N.Y.). Hybridization and the strength of hybridization (*i.e.,* the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the Tm of the formed hybrid, and the G:C ratio within the nucleic acids. As used herein, the term "Tm" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the Tm of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the Tm value may be calculated by the equation: Tm=81.5+0.41(% G+C), when a nucleic acid is in aqueous solution at 1 M NaCl [see *e.g.,* Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985)]. Other references include more sophisticated computations, which take structural as well as sequence characteristics into account for the calculation of Tm. The person skilled in the art knows well that numerous hybridization conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.,* the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high hybridization stringency Those skilled in the art know that higher stringencies are preferred to reduce or eliminate non-specific binding between the nucleotide sequence of an inventive intron and other nucleic acid sequences, whereas lower stringencies are preferred to detect a larger number of nucleic acid sequences having different homologies to the inventive nucleotide sequences. Such conditions are described by, *e.g.,* Sambrook (Molecular Cloning; A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)) or in Current Protocols in Molecular Biology, John Wiley & Sons, N. Y. (1989) 6.3.1-6.3.6. Preferred hybridization condition are disclose in the detailed description.

[0065] Identity: when used in relation to nucleic acids refers to a degree of complementarity. Identity between two nucleic acids is understood as meaning the identity of the nucleic acid sequence over in each case the entire length of the sequence, which is calculated by comparison with the aid of the program algorithm GAP (Wisconsin Package Version 10.0, University of Wisconsin, Genetics Computer Group (GCG), Madison, USA) with the parameters being set as follows:

Gap Weight: 12	Length Weight: 4
Average Match: 2,912	Average Mismatch:-2,003

[0066] For example, a sequence with at least 95% identity to the sequence SEQ ID NO. 1 at the nucleic acid level is understood as meaning the sequence that, upon comparison with the sequence SEQ ID NO. 1 by the above program algorithm with the above parameter set, has at least 95% identity. There may be partial identity (*i.e.*, partial identity of less then 100%) or complete identity (*i.e.,* complete identity of 100%).

[0067] Introducing a recombinant DNA expression construct: in plant cells refers to a recombinant DNA expression construct that will be introduced into the genome of a plant by transformation and is stably maintained. The term "introducing" encompasses for example methods such as transfection, transduction or transformation.

[0068] Identification, "Identifying" or "selecting": with regard to transformation of plants has to be understood as a

screening procedure to identify and select those plant cells in which the recombinant expression construct has been introduced stably into the genome. "Identifying" with regard to an intron with gene expression enhancing properties refers to a process for the selection of said intron out of a population of introns. Preferably, "identifying" refers to an *in silico* selection process, more preferably to an automated *in silico* selection process, using the selection criteria of the inventive methods. Such an *in silico* identification process can comprise for instance the steps of

(1) generating an intron sequence database on the basis of DNA sequences present in a DNA sequence database (e.g. genomic DNA databases publicly available via the internet),
(2) screening of the generated intron DNA sequence database -or other genomic DNA sequences containing databases - for introns with gene expression enhancing properties using the criteria according to the inventive method,

wherein the steps for retrieving or generating the DNA sequences, the generation of an intron specific DNA sequence database and the screening of these DNA sequences - using the criteria according to the inventive method - will be performed with the aid of appropriate computer algorithms and computer devices.

[0069] Intron: refers to sections of DNA (intervening sequences) within a gene that do not encode part of the protein that the gene produces, and that is spliced out of the mRNA that is transcribed from the gene before it is exported from the cell nucleus. Intron sequence refers to the nucleic acid sequence of an intron. Thus, introns are those regions of DNA sequences that are transcribed along with the coding sequence (exons) but are removed during the formation of mature mRNA. Introns can be positioned within the actual coding region or in either the 5' or 3' untranslated leaders of the pre-mRNA (unspliced mRNA). Introns in the primary transcript are excised and the coding sequences are simultaneously and precisely ligated to form the mature mRNA. The junctions of introns and exons form the splice site. The sequence of an intron begins with GU and ends with AG. Furthermore, in plants, two examples of AU-AC introns have been described: the fourteenth intron of the RecA-like protein gene and the seventh intron of the G5 gene from Arabidopsis thaliana are AT-AC introns. Pre-mRNAs containing introns have three short sequences that are -beside other sequences- essential for the intron to be accurately spliced. These sequences are the 5' splice-site, the 3' splice-site, and the branchpoint. mRNA splicing is the removal of intervening sequences (introns) present in primary mRNA transcripts and joining or ligation of exon sequences. This is also known as cis-splicing which joins two exons on the same RNA with the removal of the intervening sequence (intron). The functional elements of an intron comprising sequences that are recognized and bound by the specific protein components of the spliceosome (e.g. splicing consensus sequences at the ends of introns). The interaction of the functional elements with the spliceosome results in the removal of the intron sequence from the premature mRNA and the rejoining of the exon sequences. Introns have three short sequences that are essential -although not sufficient- for the intron to be accurately spliced. These sequences are the 5' splice site, the 3' splice site and the branch point. The branchpoint sequence is important in splicing and splice-site selection in plants. The branchpoint sequence is usually located 10-60 nucleotides upstream of the 3' splice site. Plant sequences exhibit sequence deviations in the branchpoint, the consensus sequences being 5-CURAY-3' (SEQ ID NO:75) or 5'-YURAY-3' (SEO ID NO: 76).

[0070] "IME-intron" or intron mediated enhancement (IME)-intron: when made in reference to an intron sequence refers to an intron with gene expression enhancing properties in plants as defined herein (see gene expression enhancing properties, gene expression enhancing effect or intron mediated gene expression enhancement).

[0071] Isolation or isolated: when used in relation to an intron or gene, as in "isolation of an intron sequence" or "isolation of a gene" refers to a nucleic acid sequence that is identified within and isolated/separated from its chromosomal nucleic acid sequence context within the respective source organism. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA, which are found in the state they exist in nature. For example, a given DNA sequence (e.g. a gene) is found on the host cell chromosome in proximity to neighboring genes; intron sequences, are imbedded into the nucleic acid sequence of a gene in an alternating sequence of introns and exons. The isolated nucleic acid sequence may be present in single-stranded or double-stranded form. When an isolated nucleic acid sequence is to be utilized to express a protein, the nucleic acid sequence will contain at a minimum at least a portion of the sense or coding strand (*i.e.,* the nucleic acid sequence may be single-stranded). Alternatively, it may contain both the sense and antisense strands (*i.e.,* the nucleic acid sequence may be double-stranded).

[0072] Nucleic acid: refers to deoxyribonucleotides, ribonucleotides or polymers or hybrids thereof in single- or double-stranded, sense or antisense form. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.,* degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The term "nucleic acid" can be used to describe a "gene", "cDNA","DNA" "mRNA", "oligonucleotide," and "polynucleotide".

[0073] Nucleic acid sequence: as used herein refers to the consecutive sequence of deoxyribonucleotides or ribonucleotides (nucleotides) of a DNA fragment (oligonucleotide, polynucleotide, genomic DNA, cDNA etc.) as it can made be available by DNA sequencing techniques as a list of abbreviations, letters, characters or words, which represent

nucleotides.

[0074] Organ: with respect to a plant (or "plant organ") means parts of a plant and may include (but shall not limited to) for example roots, fruits, shoots, stem, leaves, anthers, sepals, petals, pollen, seeds, etc.

[0075] Otherwise unchanged conditions: means - for example - that the expression which is initiated by one of the expression constructs to be compared is not modified by combination with additional genetic control sequences, for example enhancer sequences and is done in the same environment (*e.g.,* the same plant species) at the same developmental stage and under the same growing conditions.

[0076] Plant: is generally understood as meaning any single-or multi-celled organism or a cell, tissue, organ, part or propagation material (such as seeds or fruit) of same which is capable of photosynthesis. Included for the purpose of the invention are all genera and species of higher and lower plants of the Plant Kingdom. Annual, perennial, monocotyledonous and dicotyledonous plants are preferred. The term includes the mature plants, seed, shoots and seedlings and their derived parts, propagation material (such as seeds or microspores), plant organs, tissue, protoplasts, callus and other cultures, for example cell cultures, and any other type of plant cell grouping to give functional or structural units. Mature plants refer to plants at any desired developmental stage beyond that of the seedling. Seedling refers to a young immature plant at an early developmental stage. Annual, biennial, monocotyledonous and dicotyledonous plants are preferred host organisms for the generation of transgenic plants. The expression of genes is furthermore advantageous in all ornamental plants, useful or ornamental trees, flowers, cut flowers, shrubs or lawns. Plants which may be mentioned by way of example but not by limitation are angiosperms, bryophytes such as, for example, *Hepaticae* (liverworts) and *Musci* (mosses); *Pteridophytes* such as ferns, horsetail and club mosses; gymnosperms such as conifers, cycads, ginkgo and *Gnetatae*; algae such as *Chlorophyceae, Phaeophpyceae, Rhodophyceae, Myxophyceae, Xanthophyceae, Bacillariophyceae* (diatoms), and *Euglenophyceae.* Preferred are plants which are used for food or feed purpose such as the families of the *Leguminosae* such as pea, alfalfa and soya; *Gramineae* such as rice, maize, wheat, barley, sorghum, millet, rye, triticale, or oats; the family of the *Umbelliferae,* especially the genus *Daucus,* very especially the species *carota* (carrot) and *Apium,* very especially the species *Graveolens dulce* (celery) and many others; the family of the *Solanaceae,* especially the genus *Lycopersicon,* very especially the species *esculentum* (tomato) and the genus *Solanum,* very especially the species *tuberosum* (potato) and *melongena* (egg plant), and many others (such as tobacco); and the genus *Capsicum,* very especially the species *annuum* (peppers) and many others; the family of the *Leguminosae,* especially the genus *Glycine*, very especially the species *max* (soybean), alfalfa, pea, lucerne, beans or peanut and many others; and the family of the *Cruciferae* (*Brassica*cae), especially the genus *Brassica,* very especially the species *napus* (oil seed rape), *campestris* (beet), *oleracea* cv Tastie (cabbage), *oleracea* cv Snowball Y (cauliflower) and *oleracea* cv Emperor (broccoli); and of the genus *Arabidopsis,* very especially the species *thaliana* and many others; the family of the *Compositae,* especially the genus *Lactuca,* very especially the species *sativa* (lettuce) and many others; the family of the *Asteraceae* such as sunflower, Tagetes, lettuce or Calendula and many other; the family of the *Cucurbitaceae* such as melon, pumpkin/squash or zucchini, and linseed. Further preferred are cotton, sugar cane, hemp, flax, chillies, and the various tree, nut and wine species.

[0077] Providing: when used in relation to an intron as in "physically providing an intron" refers to the cloning of the DNA sequence representing said intron from a plant of interest and the provision of such an intron physically in an appropriate vector or plasmid for further cloning work and the subsequent application of said intron according to the invention.

[0078] Producing: when used in relation to an intron as in "producing an intron" refers to the synthesis of DNA molecules on the basis of DNA sequence information of an inventive intron.

[0079] Promoter, promoter element, or promoter sequence: as used herein, refers to a DNA sequence which when ligated to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into mRNA. Thus, a promoter is a recognition site on a DNA sequence that provide an expression control element for a gene and to which RNA polymerase specifically binds and initiates RNA synthesis (transcription) of that gene. A promoter is typically, though not necessarily, located 5' (*i.e.,* upstream) of a nucleotide sequence of interest (*e.g.,* proximal to the transcriptional start site of a structural gene). Promoters may be tissue specific or cell specific. The term "tissue specific" as it applies to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (*e.g.,* petals) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (*e.g.,* roots). Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (*e.g.,* heat shock, chemicals, light, etc.). Typically, constitutive promoters are capable of directing expression of a transgene in substantially any cell and any tissue. In contrast, a "regulatable" promoter is one which is capable of directing a level of transcription of an operably linked nuclei acid sequence in the presence of a stimulus (*e.g.,* heat shock, chemicals, light, etc.) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus. A promoter sequence functioning in plants is understood as meaning, in principle, any promoter which is capable of governing the expression of genes, in particular foreign genes, in plants or plant parts, plant cells, plant tissues or plant cultures. In this context, expression

can be, for example, constitutive, inducible or development-dependent. A constitutive promoter is a promoter where the rate of RNA polymerase binding and initiation is approximately constant and relatively independent of external stimuli. Usable promoters are constitutive promoters (Benfey et al. (1989) EMBO J. 8:2195-2202), such as those which originate from plant viruses, such as 35S CAMV (Franck et al. (1980) Cell 21:285-294), 19S CaMV (see also US 5352605 and WO 84/02913), 34S FMV (Sanger et al. (1990) Plant. Mol. Biol., 14:433-443), the parsley ubiquitin promoter, or plant promoters such as the Rubisco small subunit promoter described in US 4,962,028 or the plant promoters PRP1 [Ward et al. (1993) Plant. Mol. Biol. 22: 361-6], SSU, PGEL1, OCS [Leisner (1988) Proc Natl Acad Sci USA 85(5):2553-2557], lib4, usp, mas [Comai (1990) Plant Mol Biol 15(3):373-381], STLS1, ScBV (Schenk (1999) Plant Mol Biol 39(6): 1221-1230), B33, SAD1 or SAD2 (flax promoters, Jain et al. (1999) Crop Science 39(6) :1696-1701) or nos [Shaw et al. (1984) Nucleic Acids Res. 12(20):7831-7846]. An inducible promoter is a promoter where the rate of RNA polymerase binding and initiation is modulated by external stimuli. Such stimuli include light, heat, anaerobic stress, alteration in nutrient conditions, presence or absence of a metabolite, presence of a ligand, microbial attack, wounding and the like (for a review, see Gatz (1997) Annu. Rev. Plant Physiol. Plant Mol. Biol. 48:89-108). Chemically inducible promoters are particularly suitable when it is desired to express the gene in a time-specific manner. Examples of such promoters are a salicylic acid inducible promoter (WO 95/19443), and abscisic acid-inducible promoter (EP 335 528), a tetracycline-inducible promoter (Gatz et al. (1992) Plant J. 2:397-404), a cyclohexanol- or ethanol-inducible promoter (WO 93/21334) or others as described herein. A viral promoter is a promoter with a DNA sequence substantially similar to the promoter found at the 5' end of a viral gene. A typical viral promoter is found at the 5' end of the gene coding for the p21 protein of MMTV described by Huang et al. ((1981) Cell 27:245). A synthetic promoter is a promoter that was chemically synthesized rather than biologically derived. Usually synthetic promoters incorporate sequence changes that optimize the efficiency of RNA polymerase initiation. A temporally regulated promoter is a promoter where the rate of RNA polymerase binding and initiation is modulated at a specific time during development. Examples of temporally regulated promoters are given in Chua et al. [(1989) Science 244:174-181]. A spatially regulated promoter is a promoter where the rate of RNA polymerase binding and initiation is modulated in a specific structure of the organism such as the leaf, stem or root. Examples of spatially regulated promoters are given in Chua et al. [(1989) Science 244:174-181]. A spatiotemporally regulated promoter is a promoter where the rate of RNA polymerase binding and initiation is modulated in a specific structure of the organism at a specific time during development. A typical spatiotemporally regulated promoter is the EPSP synthase-35S promoter described by Chua et al. [(1989) Science 244:174-181]. Suitable promoters are furthermore the oilseed rape napin gene promoter (US 5,608,152), the Vicia faba USP promoter (Bäumlein et al. (1991) Mol Gen Genet 225(3):459-67), the Arabidopsis oleosin promoter (WO 98/45461), the Phaseolus vulgaris phaseolin promoter (US 5,504,200), the Brassica Bce4 promoter (WO 91/13980), the bean arc5 promoter, the carrot DcG3 promoter, or the Legumin B4 promoter (LeB4; Bäumlein et al. (1992) Plant Journal 2(2):233-9), and promoters which bring about the seed-specific expression in monocotyledonous plants such as maize, barley, wheat, rye, rice and the like. Advantageous seed-specific promoters are the sucrose binding protein promoter (WO 00/26388), the phaseolin promoter and the napin promoter. Suitable promoters which must be considered are the barley lpt2 or lpt1 gene promoter (WO 95/15389 and WO 95/23230), and the promoters described in WO 99/16890 (promoters from the barley hordein gene, the rice glutelin gene, the rice oryzin gene, the rice prolamin gene, the wheat gliadin gene, the wheat glutelin gene, the maize zein gene, the oat glutelin gene, the sorghum kasirin gene and the rye secalin gene). Further suitable promoters are Amy32b, Amy 6-6 and Aleurain [US 5,677,474], Bce4 (oilseed rape) [US 5,530,149], glycinin (soya) [EP 571 741], phosphoenolpyruvate carboxylase (soya) [JP 06/62870], ADR12-2 (soya) [WO 98/08962], isocitrate lyase (oilseed rape) [US 5,689,040] or α-amylase (barley) [EP 781 849]. Other promoters which are available for the expression of genes in plants are leaf-specific promoters such as those described in DE-A 19644478 or light-regulated promoters such as, for example, the pea petE promoter. Further suitable plant promoters are the cytosolic FBPase promoter or the potato ST-LSI promoter (Stockhaus et al. (1989) EMBO J. 8:2445), the Glycine max phosphoribosylpyrophosphate amidotransferase promoter (GenBank Accession No. U87999) or the node-specific promoter described in EP-A-0 249 676. Other suitable promoters are those which react to biotic or abiotic stress conditions, for example the pathogen-induced PRP1 gene promoter (Ward et al.. (1993) Plant. Mol. Biol. 22:361-366), the tomato heat-inducible hsp80 promoter (US 5,187,267), the potato chill-inducible alpha-amylase promoter (WO 96/12814) or the wound-inducible pinII promoter (EP-A-0 375 091) or others as described herein. Other promoters, which are particularly suitable, are those that bring about plastid-specific expression. Suitable promoters such as the viral RNA polymerase promoter are described in WO 95/16783 and WO 97/06250, and the Arabidopsis clpP promoter, which is described in WO 99/46394. Other promoters, which are used for the strong expression of heterologous sequences in as many tissues as possible, in particular also in leaves, are, in addition to several of the abovementioned viral and bacterial promoters, preferably, plant promoters of actin or ubiquitin genes such as, for example, the rice actin1 promoter. Further examples of constitutive plant promoters are the sugarbeet V-ATPase promoters (WO 01/14572). Examples of synthetic constitutive promoters are the Super promoter (WO 95/14098) and promoters derived from G-boxes (WO 94/12015). If appropriate, chemical inducible promoters may furthermore also be used, compare EP-A 388186, EP-A 335528, WO 97/06268. The above listed promoters can be comprise other regulatory elements that affect gene expression in response to plant hormones (Xu et al., 1994,

Plant Cell 6(8):1077-1085) biotic or abiotic environmental stimuli, such as stress conditions, as exemplified by drought (Tran et al. (2004) Plant Cell 16(9):2481-2498), heat, chilling, freezing, salt stress, oxidative stress (US 5,290,924) or biotic stressors like bacteria, fungi or viruses.

[0080] Polypeptide, peptide, oligopeptide, gene product, expression product and protein: are used interchangeably herein to refer to a polymer or oligomer of consecutive amino acid residues.

[0081] Recombinant or transgenic DNA expression construct: with respect to, for example, a nucleic acid sequence (expression construct, expression cassette or vector comprising said nucleic acid sequence) refers to all those constructs originating by experimental manipulations in which either

a) said nucleic acid sequence, or

b) a genetic control sequence linked operably to said nucleic acid sequence (a), for example a promoter, or

c) (a) and (b)

is not located in its natural genetic environment or has been modified by experimental manipulations, an example of a modification being a substitution, addition, deletion, inversion or insertion of one or more nucleotide residues. Natural genetic environment refers to the natural chromosomal locus in the organism of origin, or to the presence in a genomic library. In the case of a genomic library, the natural genetic environment of the nucleic acid sequence is preferably retained, at least in part. The environment flanks the nucleic acid sequence at least at one side and has a sequence of at least 50 bp, preferably at least 500 bp, especially preferably at least 1,000 bp, very especially preferably at least 5,000 bp, in length. A naturally occurring expression construct - for example the naturally occurring combination of a promoter with the corresponding gene - becomes a transgenic expression construct when it is modified by non-natural, synthetic "artificial" methods such as, for example, mutagenesis. Such methods have been described (US 5,565,350; WO 00/15815). Recombinant polypeptides or proteins: refer to polypeptides or proteins produced by recombinant DNA techniques, *i.e.,* produced from cells transformed by an exogenous recombinant DNA construct encoding the desired polypeptide or protein. Recombinant nucleic acids and polypeptide may also comprise molecules which as such does not exist in nature but are modified, changed, mutated or otherwise manipulated by man. An important use of the intron sequences of the invention will be the enhancement of the expression of a nucleic acid sequence, which encodes a particular protein, a polypeptide or DNA sequences that interfere with normal transcription or translation, e.g. interference- or antisense-RNA. In one embodiment of the present invention, the recombinant DNA expression construct confers expression of one or more nucleic acid molecules. Said recombinant DNA expression construct according to the invention advantageously encompasses a promoter functioning in plants, additional regulatory or control elements or sequences functioning in plants, an intron sequence with expression enhancing properties in plants and a terminator functioning in plants. Additionally, the recombinant expression construct might contain additional functional elements such as expression cassettes conferring expression of e.g. positive and negative selection markers, reporter genes, recombinases or endonucleases effecting the production, amplification or function of the expression cassettes, vectors or recombinant organisms according to the invention. Furthermore, the recombinant expression construct can comprise nucleic acid sequences homologous to a plant gene of interest having a sufficient length in order to induce a homologous recombination (HR) event at the locus of the gene of interest after introduction in the plant. A recombinant transgenic expression cassette of the invention (or a transgenic vector comprising said transgenic expression cassette) can be produced by means of customary recombination and cloning techniques as are described (for example, in Maniatis 1989, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor (NY); Silhavy 1984,) Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; and in Ausubel 1987, Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience). The introduction of an expression cassette according to the invention into an organism or cells, tissues, organs, parts or seeds thereof (preferably into plants or plant cells, tissue, organs, parts or seeds) can be effected advantageously using vectors, which comprise the above described nucleic acids, promoters, introns, terminators, regulatory or control elements and functional elements.

[0082] Regeneration: as used herein, means growing a whole plant from a plant cell, a group of plant cells, a plant part or a plant piece (*e.g.,* from a protoplast, callus, protocorm-like body, or tissue part).

[0083] Regulatory sequence: refers to promoters, enhancer or other segments of DNA where regulatory proteins such as transcription factors bind and thereby influencing the transcription rate of a given gene.

[0084] Substantially all introns of a plant genome represented in a genomic DNA sequence database or genomic DNA library: refers to more than 80%, preferably to more than 90%, more preferably to more than 95%, still more preferably more than 98% of all introns present in the genome of the plant used as a source for the preparation of the genomic DNA sequence database or genomic DNA library. The construction of genomic libraries and the subsequent sequencing of the genomic DNA and the construction of a genomic or genome DNA sequence database using the obtained sequence information is well established in the art (Mozo et al. (1998) Mol. Gen. Genet. 258:562-570; Choi et al. (1995) Weeds

World 2 :17-20; Lui et al. (1999) Proc. Natl. Acad. Sci. USA 96:6535-6540; The Arabidopsis Genome initiative, Nature 402:761-777 (1999); The Arabidopsis Genome initiative, Nature 408:796-826 (2000).

[0085] Structural gene: as used herein is intended to mean a DNA sequence that is transcribed into mRNA which is then translated into a sequence of amino acids characteristic of a specific polypeptide.

[0086] Sufficient length: with respect to a homology sequence comprised in a DNA-construct (*e.g.,* the homology sequence A or B) is to be understood to comprise sequences of a length of at least 100 base pair, preferably at least 250 base pair, more preferably at least 500 base pair, especially preferably at least 1,000 base pair, most preferably at least 2,500 base pair. The term "sufficient homology" with respect to a homology sequence comprised in a DNA-construct (*e.g.,* the homology sequence A or B) is to be understood to comprise sequences having a homology to the corresponding target sequence comprised in the chromosomal DNA (*e.g.,* the target sequence A' or B') of at least 70 %, preferably at least 80 %, more preferably at least 90 %, especially preferably at least 95 %, more especially preferably at least 99%, most preferably 100 %, wherein said homology extends over a length of at least 50 base pair, preferably at least 100 base pair, more preferably at least 250 base pair, most preferably at least 500 base pair.

[0087] Target region/sequence: of a nucleic acid sequence is a portion of a nucleic acid sequence that is identified to be of interest. A "coding region" of a nucleic acid sequence is the portion of the nucleic acid sequence, which is transcribed and translated in a sequence-specific manner to produce into a particular polypeptide or protein when placed under the control of appropriate regulatory sequences. The coding region is said to encode such a polypeptide or protein.

[0088] Tissue: with respect to a plant (or "plant tissue") means arrangement of multiple plant cells including differentiated and undifferentiated tissues of plants. Plant tissues may constitute part of a plant organ (*e.g.,* the epidermis of a plant leaf) but may also constitute tumor tissues and various types of cells in culture (*e.g.,* single cells, protoplasts, embryos, calli, protocorm-like bodies, etc.). Plant tissue may be *in planta,* in organ culture, tissue culture, or cell culture.

[0089] Transforming or transformation: as used herein refers to the introduction of genetic material (*e.g.,* a transgene) into a cell. Transformation of a cell may be stable or transient. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA) which detects the presence of a polypeptide encoded by one or more of the transgenes. Alternatively, transient transformation may be detected by detecting the activity of the protein (*e.g.,* β-glucuronidase) encoded by the transgene (*e.g.,* the *uidA* gene) as demonstrated herein [*e.g.,* examples 1.6 and 2.4, histochemical assay of GUS enzyme activity by staining with X-gluc which gives a blue precipitate in the presence of the GUS enzyme; and a chemiluminescent assay of GUS enzyme activity using the GUS-Light kit (Tropix)]. The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes. In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more transgenes into the genome of a cell, preferably resulting in chromosomal integration and stable heritability through meiosis. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell with nucleic acid sequences, which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction of genomic DNA of the cell to amplify transgene sequences. The term "stable transformant" refers to a cell that has stably integrated one or more transgenes into the genomic DNA. Thus, a stable transformant is distinguished from a transient transformant in that, whereas genomic DNA from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene. Transformation also includes introduction of genetic material into plant cells in the form of plant viral vectors involving extrachromosomal replication and gene expression, which may exhibit variable properties with respect to meiotic stability.

[0090] Transgenic or recombinant: when used in reference to a cell refers to a cell which contains a transgene, or whose genome has been altered by the introduction of a transgene. The term "transgenic" when used in reference to a tissue or to a plant refers to a tissue or plant, respectively, which comprises one or more cells that contain a transgene, or whose genome has been altered by the introduction of a transgene. Transgenic cells, tissues and plants may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into a target cell or integration of the transgene into a chromosome of a target cell by way of human intervention, such as by the methods described herein.

[0091] Wild-type, natural or of natural origin: means with respect to an organism, polypeptide, or nucleic acid sequence, that said organism polypeptide, or nucleic acid sequence is naturally occurring or available in at least one naturally occurring organism polypeptide, or nucleic acid sequence which is not changed, mutated, or otherwise manipulated by man.

[0092] Vector: is a DNA molecule capable of replication in a host cell. Plasmids and cosmids are exemplary vectors. Furthermore, the terms "vector" and "vehicle" are used interchangeably in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another, whereby the cells not necessarily belonging to the same organism (e.g. transfer of a DNA segment form an *Agrobacterium* cell to a plant cell).

[0093] The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence

in a particular host organism.

DETAILED DESCRIPTION OF THE INVENTION

[0094] The teaching of the present invention enables the identification of introns causing intron mediated enhancement (IME) of gene expression. Furthermore, the present invention provides isolated plant introns that, if functionally combined with a promoter functioning in plants and a nucleic acid fragment, can enhance the expression rate of said nucleic acid in a plant or a plant cell.

[0095] A first embodiment of the present invention relates to a method for identifying an intron with plant gene expression enhancing properties comprising selecting an intron from a plant genome, wherein said intron is characterized by at least the following features

I) an intron length shorter than 1,000 base pairs, and
II) presence of a 5' splice site comprising the dinucleotide sequence 5'-GT-3' (SEQ ID NO: 78), and
III) presence of a 3' splice site comprising the trinucleotide sequence 5'-CAG-3' (SEQ ID NO: 79), and
IV) presence of a branch point resembling the consensus sequence 5'-CURAY-3' (SEQ ID NO:75) upstream of the 3'splice site, and
V) an adenine plus thymine content of at least 40% over 100 nucleotides downstream from the 5' splice site, and
VI) an adenine plus thymine content of at least 50% over 100 nucleotides upstream from the 3' splice site, and
VII) an adenine plus thymine content of at least 50%, and a thymine content of at least 30% over the entire intron.

[0096] In another embodiment, the invention relates to a method for enriching the number of introns with expression enhancing properties in plants in a population of plant introns to a percentage of at least 50% of said population, said method comprising selecting introns from said population, said introns are characterized by at least the following features

I) an intron length shorter than 1,000 base pairs, and
II) presence of a 5' splice site comprising the dinucleotide sequence 5'-GT-3' (SEQ ID NO: 78), and
III) presence of a 3' splice site comprising the trinucleotide sequence 5'-CAG-3' (SEQ ID NO: 79), and
IV) presence of a branch point resembling the consensus sequence 5'-CURAY-3' (SEQ ID NO:75) upstream of the 3'splice site, and
V) an adenine plus thymine content of at least 40% over 100 nucleotides downstream from the 5' splice site, and
VI) an adenine plus thymine content of at least 50% over 100 nucleotides upstream from the 3' splice site, and
VII) an adenine plus thymine content of at least 50%, and a thymine content of at least 30% over the entire intron.

[0097] The inclusion of any of the inventive introns described by SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 into the 5' untranslated region (UTR) of the β-glucuronidase gene (GUS) driven by the *Zea mays* Ubiquitin promoter has led to strong expression enhancement of the reporter gene in maize protoplasts (Black Mexican Sweet) suspension cells and stable transformed plants (see examples). Furthermore, it could be shown that the gene expression enhancement properties of said introns are comparable to those known from the literature (e.g. the first intron of the *Zea mays* Ubiquitin gene, used as positive control in the expression assays).

[0098] In a preferred embodiment, the number of introns - with gene expression enhancing properties - identified within a population of introns by applying the method of the invention for enrichment is enriched to a percentage of at least 50%, preferably at least 55%, more preferably at least 60%, especially preferably at least 65%, or very especially preferably at least 70% (*i.e.,* a given population of 100 introns pre-selected by using the inventive method will comprise at least 50, preferably at least 55, more preferably at least 60, especially preferably at least 65 or 70 introns with gene expression enhancing properties). More preferably, the number of introns - with gene expression enhancing properties - identified within a population of introns by applying the method of the invention for enrichment is enriched to a percentage of at least 50%, wherein the selected introns, if part of an recombinant DNA expression construct leads to an increase in the gene expression of a given gene of at least 300% compared to the otherwise identical expression construct lacking the intron under otherwise unchanged conditions. Most preferably, the enrichment is at least 60% percent, wherein the selected introns, increasing the transcription of a gene driven by a given promoter of at least 200%.Especially preferably, the enrichment is at least 70%, wherein the selected introns, increasing the transcription of a gene driven by a given promoter of at least 50%.

[0099] Preferably, the length of an inventive IME-intron is preferably shorter than 1,000 base pairs, more preferably shorter than 900 bp, most preferably shorter than 800 bp. In a preferred embodiment, the branchpoint sequence of the intron identified by a method of the invention is described by the nucleotide sequences 5'-CURAY-3' (SEQ ID NO. 75) or 5'-YURAY-3' (SEQ ID NO. 76), wherein the U and A are essential nucleotides, and purines and pyrimidines are preferred nucleotides at positions 3 and 5 respectively. In position 1, pyrimidines are preferred but also C is preferred

to U. The sequence context of the 5' splice-site surrounding the GT dinucleotide may vary. Preferred are 5' splice-sites of the sequence 5'-RR/GT(RT)(RT)(GY)-3' (SEQ ID NO. 77), wherein R stands for the nucleotides G or A, Y stands for the nucleotides C or T. The nucleotides given in brackets describing alternative nucleotides at the respective position.

[0100] In a preferred embodiment of the invention, the adenine/thymine (AT) content of an inventive intron over the entire sequence is at least 50%, more preferably at least 55%, even more preferably at least 60%.

[0101] In a preferred embodiment of the invention the populations of plant introns to which the inventive methods will be applied comprises a) substantially all introns of a plant genome represented in a DNA sequence database or b) a plant genomic DNA library. In an additional embodiment of the invention, the population of introns to which the inventive methods will be applied to is selected from the group consisting of a) introns located between two protein encoding exons, and b) introns located within the 5' untranslated region of the corresponding gene. In order to identify an intron with expression enhancing properties in plants or plant cells located within a coding region (between two protein encoding exons) or in the 5'untranslated region of a given gene, the coding regions and the 5' untranslated regions from a set of genes (*e.g.,* present in a sequence database) can be screened for the presence of introns located in said regions and the identified introns are subsequently screened using one of the inventive methods. Such an *in silico* identification process using bioinformatics tools known to the persons skilled in the art can be performed by screening a) specific DNA sequence databases (*e.g.,* containing solely coding regions or the 5' untranslated regions), or b) other publicly accessible genomic DNA sequences containing databases. In a preferred embodiment of the invention, the introns with expression enhancing properties located in the 5'untranslated regions are identified by a method comprising the steps of:

a. identifying a coding sequences within a set of genes present in a sequence database, and

b. identifying EST sequences corresponding to the genes identified under (a), and

c. comparing said coding sequences and EST sequences with the genomic sequence of the respective genes, and

d. selecting EST sequences comprising the 5' untranslated region, and

e. identifying introns located in said 5' untranslated regions.

[0102] Preferably, the steps of retrieving or generating DNA sequences or the generation of specific DNA sequence database and screening the same (e.g. using the criteria according to the inventive methods) can be performed with the aid of appropriate bioinformatic computer algorithms and appropriate computer devices known to a skilled person. In a preferred embodiment, the introns where selected from a population of introns derived from monocotyledonous plants, especially preferred are monocotyledonous plants selected from the group consisting of the genera *Hordeum, Avena, Secale, Triticum, Sorghum, Zea, Saccharum* and *Oryza.*

[0103] In a furthermore preferred embodiment of the invention, the population of introns to which the inventive methods will be applied are selected from a population of plant genes representing the 10% fraction (9^{th} decile) of genes with the highest expression rate in a gene expression analysis experiment performed using a plant cell, plant tissue or a whole plant.

[0104] To allow the determination of gene expression levels, a number of different techniques have been proposed (Milosavljevic, A. et al. (1996) Genome Res. 6:132-141; Shoemaker, D. et al. (1996) Nature Genet. 14:450-456; Sikela, J.M. and Auffray,C. (1993) Nature Genet. 3:189-191; Meier-Ewert S. et al. (1998) Nucleic Acids Research 26(9): 2216-2223). Therefore, a number of different gene expression analysis systems could be employed in accordance with the instant invention, including, but not limited to microarray analysis, "digital northern", clone distribution analysis of cDNA libraries using the "DNA sequencing by hybridization method" (Strezoska, Z. et al. (1991) Proc. Natl. Acad. Sci. USA 88:10089-10093) and Serial Analysis of Gene Expression (SAGE, Velculescu, V. E. et al. (1995) Science 270: 484-487).

[0105] By using the cDNA microarray hybridization technology the expression profiles of thousands of genes can be monitored at once. The DNA array analysis has become a standard technique in the molecular biology laboratory for monitoring gene expression. Arrays can be made either by the mechanical spotting of pre-synthesized DNA products or by the *de novo* synthesis of oligonucleotides on a solid substrate, usually a derivatized glass slide. Typically arrays are used to detect the presence of mRNAs that may have been transcribed from different genes and which encode different proteins. The RNA is extracted from many cells, or from a single cell type, then converted to cDNA or cRNA. The copies may be "amplified" by (RT-) PCR. Fluorescent tags are enzymatically incorporated into the newly synthesized strands or can be chemically attached to the new strands of DNA or RNA. A cDNA or cRNA molecule that contains a sequence complementary to one of the single-stranded probe sequences will hybridize, or stick, via base pairing to the spot at which the complementary probes are affixed. The spot will then fluoresce when examined using a microarray scanner. Increased or decreased fluorescence intensity indicates that cells in the sample have recently transcribed, or

ceased transcription, of a gene that contains the probed sequence. The intensity of the fluorescence is proportional to the number of copies of a particular mRNA that were present and thus roughly indicates the activity or expression level of that gene. Microarrys (and the respective equipment needed to perform the expression analysis experiments) that can be employed in accordance with the present invention are commercially available. The GeneChip Arabidopsis ATH1 Genome Array, produced from Affimetrix (Santa Clara, CA), contains more than 22,500 probe sets representing approximately 24,000 genes. The array is based on information from the international Arabidopsis sequencing project that was formally completed in December 2000 (http://www.affymetrix. com). Thus, the expression rate of the analyzed genes can be ranked (according to the intensity of the fluorescence of the respective genes after the hybridization process) and the genes belonging to the 10% of genes showing the highest gene expression rate can be identified by using microarray analysis.

[0106] Databases containing microarray expression profiling results are publicly available via the internet e.g. the Nottingham Arabidopsis Stock Center's microarray database or the OSMID (osmotic stress microarray information) database. The Nottingham *Arabidopsis* Stock Center's microarray database containing a wide selection of microarray data from Affimetrix gene chips (http://affymetrix.arabidopsis. info). The OSMID database (http://www.osmid.org) contains the results of approximately 100 microarray experiments performed at the University of Arizona. This includes analysis of NaCl, cold, and drought treatments of *Arabidopsis thaliana,* rice (*Oryza sativa*), barley, (*Hordeum vulgaris*), ice plant (*Mesembryanthemum crystallinum*), and corn (*Zea mays*). Thus, by using the expression profiles present in sequence/expression databases the expression rate of genes can be ranked (according to the clone distribution of the respective cDNA in the library) and genes belonging to the 10% of genes showing the highest (abundance) gene expression rate can be identified.

[0107] "Digital Northern" are generated by partially sequencing thousands of randomly selected clones from relevant cDNA libraries. Differentially expressed genes can then be detected from variations in the counts of their cognate sequence tags. The sequence tag-based method consists of generating a large number (thousands) of expressed sequence tags (ESTs) from 3'-directed regional non-normalized cDNA libraries. The concept of a "digital Northern" comparison is the following: a number of tags is reported to be proportional to the abundance of cognate transcripts in the tissue or cell type used to make the cDNA library. The variation in the relative frequency of those tags, stored in computer databases, is then used to point out the differential expression of the corresponding genes (Okubo *et al.* 1992; Matsubara and Okubo 1994). The SAGE method is a further development of this technique, which requires only nine nucleotides as a tag, therefore allowing a larger throughput. Thus, the expression rate of the analyzed genes by using the "digital Northern" method can be ranked (according to the abundance of the tags of the respective gene in the cDNA library) and the genes belonging to the 10% of genes showing the highest (abundance) gene expression rate can be identified.

[0108] Using the "sequencing by hybridization method" described in the US patents US 5,667,972, US 5,492,806, US 5,695,940, US 5,972,619, US 6,018,041, US 6,451,996, US 6,309,824 it is possible to perform *in silico* clone distribution analysis of complete cDNA libraries. The entire content of said US patents is incorporated by reference. This technology is commercially available and customized experiments can be conducted in collaboration with the company HySeq Inc.. To determine clone distribution by using the "sequencing by hybridization method", or "HySeq-technology" plants are grown under a variety of conditions and treatments, and then tissues at different developmental stages are collected. This is done in a strategic manner so the probability of harvesting all expressible genes in at least one or more of the libraries is maximized. mRNA is then extracted from each of the collected samples and used for the library production. The libraries can be generated from mRNA purified on oligo dT columns. Colonies from transformation of the cDNA library into *E.*coli are randomly picked and placed into microtiter plates and subsequently spotted DNA onto a surface. The cDNA inserts from each clone from the microtiter plates are PCR amplified and spotted onto a nylon membrane. A battery of 288 ^{33}-P radiolabeled seven-mer oligonucleotides are then sequentially hybridized to the membranes. After each hybridization a blot image is captured during a phosphorimage scan to generate a profile for each single oligonucleotide. Absolute identity is maintained by barcoding for image cassette, filter and orientation within the cassette. The filters are then treated using relatively mild conditions to strip the bound probes and then returned to the hybridization chambers for another round. The hybridization and imaging cycle is repeated until the set of 288 oligomers is completed. After completion of hybridizations, each spot (representing a cDNA insert) will have recorded the amount of radio signal generated from each of the 288 seven-mer oligonucleotides. The profile of which oligomers bound, and to what degree, to each single cDNA insert (a spot on the membrane) is defined as the *signature* generated from that clone. Each clone's signature is compared with all other signatures generated from the same organism to identify clusters of related signatures. This process "sorts" all of the clones from an organism into so called "clusters" before sequencing. In the clustering process, complex or tissue specific cDNA libraries are "mined" using a series of 288 seven base-pair oligonucleotides. By collecting data on the hybridization signature of these oligos, the random set of clones in a library can be sorted into "clusters". A cluster is indicative for the abundance of each gene in a particular library and is therefore a measure of the gene expression rate of an individual gene. Thus, the expression rate of genes can be ranked using the "HySeq" technology and the genes belonging to the 10% of genes showing the highest (abundance) gene expression rate can

be identified.

[0109] The genes, cDNAs or expressed sequence tags chosen for the identification of the inventive introns, belonging to the 10%, preferably 5%, more preferably 3% most preferably 1% of genes showing the highest gene expression rate in a gene expression analysis experiment, wherein the gene expression rate can be calculated indirectly by using the above described methods. In a preferred embodiment of the invention, the nucleic acid sequences of the genes belonging to the 10% of genes showing the highest gene expression rate where used to isolate the complete genomic DNA sequence - including the intron sequences- of the respective genes by screening of e.g. appropriate DNA sequences containing databases, or genomic DNA or genomic DNA libraries using hybridization methods or RACE cloning techniques (rapid amplification of cDNA ends), or chromosome walking techniques. After sequence determination of the isolated complete genomic DNA of the respective candidate gene, the intron sequences present in said genes were screened using the above described criteria to identify those introns, having expression enhancing properties. The described *in silico* methods for the selection of introns with expression enhancing properties have a high probability of success, but the efficiency of the described methods may be further increased by combination with other methods. Therefore, in one preferred embodiment of the invention independent validation of the genes representing the 10% of genes showing the highest gene expression rate in a gene expression analysis experiment is done using alternative gene expression analysis tools, like Northern analysis, or real time PCR analysis (see examples).

[0110] In a preferred embodiment of the invention the method for the identification or enrichment of introns with gene expression enhancing properties in plants is applied to DNA sequence databases using an automated process, more preferably using a computer device and an algorithm that defines the instructions needed for accomplishing the selection steps for identifying or enriching introns with gene expression enhancing properties in plants within the screened population of DNA sequences. A further embodiment of the invention is a computer algorithm that defines the instructions needed for accomplishing the selection steps for identifying or enriching introns with plant gene expression enhancing properties as described above. Useful computer algorithms are well known in the art of bioinformatics or computational biology. Bioinformatics or computational biology is the use of mathematical and informational techniques to analyze sequence data (e.g. generation of sequence data, sequence alignments, screening of sequence data) usually by creating or using computer programs, mathematical models or both. One of the main areas of bioinformatics is the data mining and analysis of data gathered from different sources. Other areas are sequence alignment, protein structure prediction. Another aspect of bioinformatics in sequence analysis is the automatic search for genes or regulatory sequences within a genome (e.g. intron sequences within a stretch of genomic DNA). Sequence databases can be searched using a variety of methods. The most common is probably searching for a sequence similar to a certain target gene whose sequence is already known to the user. A useful program is the BLAST (Basic Local Alignment Search Tool) program a method of this type. BLAST is an algorithm for comparing biological sequences, such as DNA sequences of different genes. Given a library or database of sequences, a BLAST search enables a researcher to look for specific sequences. The BLAST algorithm and a computer program that implements it were developed by Stephen Altschul at the U.S. National Center for Biotechnology Information (NCBI) and is available on the web at http://www.ncbi.nlm.nih.gov/BLAST. The BLAST program can either be downloaded and run as a command-line utility "blastall" or accessed for free over the web. The BLAST web server, hosted by the NCBI, allows anyone with a web browser to perform similarity searches against constantly updated databases of proteins and DNA that include most of the newly sequenced organisms. BLAST is actually a family of programs (all included in the blastall executable) including beside others the Nucleotide-nucleotide BLAST (BLASTN). This program, given a DNA query, returns the most similar DNA sequences from the DNA database that the user specifies. A person skilled in the art knows how to produce or retrieve sequence Data from e.g. public sequence database and to design algorithms to screen the set of sequences in a customized way (see examples).

[0111] Additionally, the invention relates to computer algorithm that defines the instructions needed for accomplishing the selection steps for identifying or enriching introns with gene expression enhancing properties in plants from a plant genome or a population of introns selected from the group consisting of introns located between two protein encoding exons, and/or introns located within the 5' untranslated region of the corresponding gene and/or introns located in the DNA sequences of genes representing the 10% fraction of genes with the highest expression rate in a gene expression analysis experiment performed using a plant cell, plant tissue and/or a whole plant. Another embodiment of the invention is a computer device or data storage device comprising the algorithm. A storage device can be a hard disc" (or "hard drive") or an optical data storage medium like a CD-ROM ("Compact Disc Read-Only Memory" (ROM) or DVD (digital versatile disc) or any other mechanically, magnetically, or optically data storage medium.

[0112] Another embodiment of the invention relates to a method for isolating, providing or producing an intron with gene expression enhancing properties in plants comprising the steps of

a) performing an identification or enrichment of introns with gene expression enhancing properties in plants as described above and providing the sequence information of said identified or enriched introns, and

b) providing the physical nucleotide sequence of said introns identified or enriched under a) and

c) evaluating the gene expression enhancing properties of the intron sequence provided under b) in an *in vivo* or

in vitro expression experiment, and
d) isolating introns from said expression experiment c), which demonstrate expression enhancing properties.

[0113] Preferably, evaluation of the gene expression enhancing properties of the isolated introns comprises,

c1) providing a recombinant expression constructs by functionally linking an individual nucleotide sequence from step b) with at least one promoter sequence functioning in plants or plant cells, and at least one readily quantifiable nucleic acid sequence, and

c2) introducing said recombinant DNA expression construct in plant cells and evaluating the gene expression enhancing properties of the isolated intron.

[0114] Preferably, the evaluation of the gene expression enhancing properties is done in a plant cell or stable transformed plants and wherein said isolated intron enhances expression of a given gene at least twofold (see examples).

[0115] An additional subject matter of the invention relates to a recombinant DNA expression construct comprising at least one promoter sequence functioning in plants cells, at least one nucleic acid sequence and at least one intron selected from the group consisting of the sequences described by SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22, and functional equivalents thereof, wherein said promoter sequence and at least one of said intron sequences are functionally linked to said nucleic acid sequence and wherein said intron is heterologous to said nucleic acid sequence or to said promoter sequence. Furthermore, the invention relates to recombinant expression constructs comprising at least one promoter sequence functioning in plants cells, at least one nucleic acid sequence and at least one functional equivalents of an intron described by any of sequences SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22.

[0116] Preferably, said functional equivalents comprising the functional elements of an intron, wherein said promoter sequence and at least one of said intron sequences are functionally linked to said nucleic acid sequence and wherein said intron is heterologous to said nucleic acid sequence or to said promoter sequence. More preferably, the functional equivalent is further characterized by

i) having at least 50 consecutive base pairs of the intron sequence described by any of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22, or
ii) having an identity of at least 80% over a sequence of at least 95 consecutive nucleic acid base pairs to a sequences described by any of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 or
iii) hybridizing under high stringent conditions with a nucleic acid fragment of at least 50 consecutive base pairs of a nucleic acid molecule described by any of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22,

[0117] In a preferred embodiment of the invention, the introns comprising at least 50 bases pairs, more preferably at least 40 bases pairs, most preferably 30 bases pairs of the sequences/exons 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively. In another embodiment of the in, the recombinant DNA expression construct of the invention further comprises one or more additional regulatory sequences functionally linked to a promoter. Those regulatory sequences can be selected from the group consisting of heat shock-, anaerobic responsive-, pathogen responsive-, drought responsive-, low temperature responsive-, ABA responsive-elements, 5'-untranslated gene region, 3'-untranslated gene region, transcription terminators, polyadenylation signals and enhancers. *Cis*- and *trans*-acting factors involved in ABA-induced gene expression have been reviewed by Bray (1997) Trends Plant Sci. 2:48-54; Busk et al. (1998) Plant Mol. Biol. 37:425-435 and Shinozaki and Yamaguchi-Shinozaki (2000) Curr. Opin. Plant Biol. 3:217-223). Many ABA-inducible genes contain a conserved, ABA-responsive, *cis*-acting element named ABRE (ABA-responsive element; PyACGTGGC) in their promoter regions (Guiltinan et al. (1990) Science 250 :267-271; Mundy et al. (1990) Proc. Natl. Acad. Sci. USA 87:406-410). The promoter region of the *rd29A* gene was analyzed, and a novel *cis*-acting element responsible for dehydration- and cold-induced expression was identified at the nucleotide sequence (Yamaguchi-Shinozaki and Shinozaki (1994) Plant Cell 6:251-264.). A 9-bp conserved sequence, TACCGACAT, termed the dehydration-responsive element (DRE), is essential for the regulation of dehydration responsive gene expression. DRE-related motifs have been reported in the promoter regions of cold- and drought-inducible genes such as *kin1, cor6.6,* and *rd17* (Wang et al. (1995) Plant Mol. Biol. 28:605-617; Iwasaki et al. (1997) Plant Physiol. 115:1287). The thermoinducibility of the heat shock genes is attributed to activation of heat shock factors (HSF). HSF act through a highly conserved heat shock promoter element (HSE) that has been defined as adjacent and inverse repeats of the motif 5'-nGAAn-3' (Amin et al. (1988) Mol Cell Biol 8:3761-3769). Examples for defense or pathogen response elements are the W-box (TTGACY) and W-box-like elements, representing binding sites for plant-specific WRKY transcription factors involved in plant development and plant responses to environmental stresses (Eulgem et al. (2000) Trends Plant Sci 5:

199-206; Robatzek S et al. (2001) Plant J 28:123-133), and the Myc-element (CACATG) (Rushton PJ et al. (1998) Curr Opin Plant Biol 1:311-315). Such regulatory sequences or elements that can be employed in conjunction with a described promoter, encompass the 5'-untranslated regions, enhancer sequences and plant polyadenylation signals. Examples of translation enhancers, which may be mentioned, are the tobacco mosaic virus 5' leader sequence (Gallie et al. (1987) Nucl Acids Res 15:8693-8711), the enhancer from the octopine synthase gene and the like. Furthermore, they may promote tissue specificity (Rouster J et al. (1998) Plant J 15:435-440). The recombinant DNA expression construct will typically include the gene of interest along with a 3' end nucleic acid sequence that acts as a signal to terminate transcription and subsequent polyadenylation of the RNA. Preferred plant polyadenylation signals are those, which essentially correspond to T-DNA polyadenylation signals from *Agrobacterium tumefaciens,* in particular gene 3 of the T-DNA (octopine synthase) of the Ti plasmid pTiACHS (Gielen et al. (1984) EMBO J 3:835-46) or functional equivalents thereof. Examples of terminator sequences, which are especially suitable, are the OCS (octopin synthase) terminator and the NOS (nopaline synthase) terminator. An expression cassette and the vectors derived from it may comprise further functional elements. The term functional element is to be understood in the broad sense and refers to all those elements, which have an effect on the generation, amplification or function of the expression cassettes, vectors or recombinant organisms according to the invention. The following may be mentioned by way of example, but not by limitation:

1. Selection markers

[0118] Selection markers are useful to select and separate successfully transformed or homologous recombined cells. To select cells which have successfully undergone homologous recombination, or else to select transformed cells, it is, also typically necessary to introduce a selectable marker, which confers resistance to a biocide (for example herbicide), a metabolism inhibitor such as 2-deoxyglucose-6-phosphate (WO 98/45456) or an antibiotic to the cells which have successfully undergone recombination. The selection marker permits the selection of the transformed cells from untransformed ones (McCormick et al. (1986) Plant Cell Reports 5:81-84).

1.1 Negative selection markers

[0119] Selection markers confer a resistance to a biocidal compound such as a metabolic inhibitor (*e.g.,* 2-deoxyglucose-6-phosphate, WO 98/45456), antibiotics (*e.g.,* kanamycin, G 418, bleomycin or hygromycin) or herbicides (*e.g.,* phosphinothricin or glyphosate). Especially preferred negative selection markers are those which confer resistance to herbicides. Examples which may be mentioned are:

- Phosphinothricin acetyltransferases (PAT; also named Bialophos resistance; bar; de Block et al. (1987) EMBO J 6: 2513-2518)
- 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) confer- ring resistance to Glyphosate (N-(phosphonomethyl)glycine),
- Glyphosate degrading enzymes (Glyphosate oxidoreductase; gox),
- Dalapon inactivating dehalogenases (deh)
- sulfonylurea- and imidazolinone-inactivating acetolactate synthases (for example mutated ALS variants with, for example, the S4 and/or Hra mutation)
- Bromoxynil degrading nitrilases (bxn)
- Kanamycin- or G418- resistance genes (NPTII; NPTI) coding *e.g.,* for neomycin phosphotransferases,
- 2-Desoxyglucose-6-phosphate phosphatase (DOGR1-Gene product; WO 98/45456; EP 0 807 836) conferring resistance against 2-desoxyglucose (Randez-Gil et al., 1995 Yeast 11:1233-1240).

Additional suitable negative selection marker are the *aadA* gene, which confers resistance to the antibiotic spectinomycin, the streptomycin phosphotransferase (SPT) gene, which allows resistance to streptomycin and the hygromycin phosphotransferase (HPT) gene, which mediates resistance to hygromycin. Especially preferred are negative selection markers that confer resistance against the toxic effects imposed by D-amino acids like *e.g.,* D-alanine and D-serine (WO 03/060133; Erikson 2004). Especially preferred as negative selection marker in this contest are the daol gene (EC: 1.4. 3.3 : GenBank Acc.-No.: U60066) from the yeast *Rhodotorula gracilis* (*Rhodosporidium toruloides*) and the *E. coli* gene *dsdA* (D-serine dehydratase (D-serine deaminase) [EC: 4.3. 1.18; GenBank Acc.-No.: J01603).

1.2) Counter selection marker

[0120] Counter selection markers are especially suitable to select organisms with defined deleted sequences comprising said marker (Koprek T et al. (1999) Plant J 19(6): 719-726). Examples for counter selection marker comprise thymidin kinases (TK), cytosine deaminases (Gleave AP et al. (1999) Plant Mol Biol. 40(2):223-35; Perera RJ et al.

(1993) Plant Mol. Biol 23(4): 793-799; Stougaard J. (1993) Plant J 3:755-761), cytochrom P450 proteins (Koprek et al. (1999) Plant J 16:719-726), haloalkandehalogenases (Naested H (1999) Plant J 18:571-576), iaaH gene products (Sundaresan V et al. (1995) Genes & Development 9:1797-1810), cytosine deaminase codA (Schlaman HRM and Hooykaas PJJ (1997) Plant J 11:1377-1385), or *tms2* gene products (Fedoroff NV & Smith DL, 1993, Plant J 3:273- 289).

1.3 Positive selection marker

[0121] Furthermore, positive selection marker can be employed. Genes like isopentenyltransferase from *Agrobacterium* tumefaciens (strain:PO22; Genbank Acc.-No.: AB025109) may - as a key enzyme of the cytokinin biosynthesis - facilitate regeneration of transformed plants (*e.g.,* by selection on cytokinin-free medium). Corresponding selection methods are described (Ebinuma 2000a,b). Additional positive selection markers, which confer a growth advantage to a transformed plant in comparison with a non-transformed one, are described *e.g.,* in EP-A 0 601 092. Growth stimulation selection markers may include (but shall not be limited to) β-Glucuronidase (in combination with *e.g.,* a cytokinin glucuronide), mannose-6-phosphate isomerase (in combination with mannose), UDP-galactose-4-epimerase (in combination with *e.g.,* galactose), wherein mannose-6-phosphate isomerase in combination with mannose is especially preferred.

2) Reporter genes

[0122] Reporter genes encode readily quantifiable proteins and, via their color or enzyme activity, make possible an assessment of the transformation efficacy, the site of expression or the time of expression. Very especially preferred in this context are genes encoding reporter proteins (Schenborn E and Groskreutz D. (1999) Mol Biotechnol. 13(1):29-44) such as the green fluorescent protein (GFP) (Sheen et al. (1995) Plant Journal 8(5):777-784; Haseloff et al. (1997) Proc Natl Acad Sci USA 94(6):2122-2127; Reichel et al. (1996) Proc Natl Acad Sci USA 93(12):5888-5893; Tian et al. (1997) Plant Cell Rep 16:267-271;WO 97/41228; Chui WL et al. (1996) Curr Biol 6:325-330; Leffel SM et al. (1997) Biotechniques. 23(5):912-8), chloramphenicol-transferase, a luciferase (Ow et al. (1986) Science 234:856-859; Millar et al. (1992) Plant Mol Biol Rep 10:324-414), the aequorin gene (Prasher et al. (1985) Biochem Biophys Res Commun 126(3):1259-1268), β-galactosidase, R locus gene (encoding a protein which regulates the production of anthocyanin pigments (red coloring) in plant tissue and thus makes possible the direct analysis of the promoter activity without addition of further auxiliary substances or chromogenic substrates; Dellaporta et al. (1988) In: Chromosome Structure and Function: Impact of New Concepts, 18th Stadler Genetics Symposium 11:263-282), with β-glucuronidase being very especially preferred (Jefferson et al. (1987) EMBO J. 6:3901-3907).

[0123] 3) Origins of replication, which ensure amplification of the expression cassettes or vectors according to the invention in, for example, *E. coli.* Examples which may be mentioned are ORI (origin of DNA replication), the pBR322 ori or the P15A ori (Sambrook et al.: Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

[0124] 4) Elements which are necessary for *Agrobacterium*-mediated plant transformation, such as, for example, the right or left border of the T-DNA or the *vir* region.

[0125] The inventive recombinant expression construct contains expressible nucleic acid sequences in addition to, or other than, nucleic acid sequences encoding for marker proteins. In a preferred embodiment of the invention the recombinant DNA expression construct comprises an nucleic acid sequence encodes for i) a protein or ii) a sense, antisense, or double-stranded RNA sequence. In a further preferred embodiment of the present invention, the recombinant DNA expression construct contains a nucleic acid sequence encoding a protein. In yet another embodiment of the invention the recombinant DNA expression construct may contain a DNA for the purpose of expressing RNA transcripts that function to affect plant phenotype without being translated into a protein. Such non protein expressing sequences comprising antisense RNA molecules, sense RNA molecules, RNA molecules with ribozyme activity, double strand forming RNA molecules (RNAi). The transgenic expression constructs of the invention can be employed for suppressing or reducing expression of endogenous target genes by "gene silencing". The skilled worker knows preferred genes or proteins whose suppression brings about an advantageous phenotype. Examples may include but are not limited to down-regulation of the β-subunit of *Arabidopsis* G protein for increasing root mass (Ullah et al. (2003) Plant Cell 15: 393-409), inactivating cyclic nucleotide-gated ion channel (CNGC) for improving disease resistance (WO 2001007596), and down-regulation of 4-coumarate-CoA ligase (4CL) gene for altering lignin and cellulose contents (US 2002138870). In yet another preferred embodiment of the invention, the transgenic expression constructs of the invention contain nucleic acids, which when transcribed, produce RNA enzymes (Ribozymes) which can act as endonucleases and catalyze the cleavage of RNA molecules with selected sequences. The cleavage of the selected RNA can result in the reduced production of their encoded polypeptide products. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Ceck 1987, Proc. Natl. Acad. Sci. USA, 84:8788-8792; Gerlach et al., 1987, Nature, 328:802-805; Forster and Symons, 1987, Cell, 49:211-220). Several different ribozyme motifs have been described with RNA cleavage activity (Symons, 1992, Annu. Rev. Biochem., 61: 641-671). Examples include sequences from group 1 self splicing

introns including Tobacco Ringspot Virus (Prody et al., 1986, Science, 231:1577-1580). Other suitable ribozymes include sequences from RNaseP with cleavage activity (Yan et al. (1992) Proc. Natl. Acad. Sci. USA 87:4144-4148), hairpin ribozyme structures (Berzal-Herranz et al. (1992) Genes and Devel. 98:1207-1210) and Hepatitis Delta virus based ribozyme (U.S. Pat. No. 5,625,047). The general design and optimization of ribozymes directed RNA cleavage activity has been discussed on detail (Haseloff and Gerlach (1988) Nature 224:585-591; Symons (1992) Annu. Rev. Biochem. 61: 641-671). The choice of a particular nucleic acid sequence to be delivered to a host cell or plant depends on the aim of the transformation. In general, the main goal of producing transgenic plants is to add some beneficial traits to the plant.

[0126] In another embodiment of the invention, the recombinant expression construct comprises a nucleic acid sequence encoding for a selectable marker protein, a screenable marker protein, a anabolic active protein, a catabolic active protein, a biotic or abiotic stress resistance protein, a male sterility protein or a protein affecting plant agronomic characteristics. Such traits include, but are not limited to, herbicide resistance or tolerance, insect resistance or tolerance, disease resistance or tolerance (viral, bacterial, fungal, nematode); stress tolerance, as exemplified by tolerance to drought, heat, chilling, freezing, salt stress, oxidative stress; increased yield, food content, male sterility, starch quantity and quality, oil content and quality, vitamin content and quality (e.g. vitamin E) and the like. One may desire to incorporate one or more nucleic acid sequences conferring any of such desirable traits. Furthermore, the recombinant expression constructs of the invention can comprise artificial transcription factors (e.g. of the zinc finger protein type; Beerli (2000) Proc Natl Acad Sci USA 97(4):1495-500). These factors attach to the regulatory regions of the endogenous genes to be expressed or to be repressed and, depending on the design of the factor, bring about expression or repression of the endogenous gene. The following may be mentioned by way of example but not by way of limitation as nucleic acid sequences or polypeptides which can be used for these applications:

[0127] Improved protection of the plant embryo against abiotic stresses such as drought, high or low temperatures, for example by overexpressing the antifreeze polypeptides from *Myoxocephalus scorpius* (WO 00/00512), *Myoxocephalus octodecemspinosus,* the *Arabidopsis thaliana* transcription activator CBF1, glutamate dehydrogenases (WO 97/12983, WO 98/11240), a late embryogenesis gene (LEA), for example from barley (WO 97/13843), calcium-dependent protein kinase genes (WO 98/26045), calcineurins (WO 99/05902), farnesyl transferases (WO 99/06580, Pei 1998), ferritin (Deak 1999), oxalate oxidase (WO 99/04013; Dunwell 1998), DREB1A factor (dehydration response element B 1A; Kasuga 1999), mannitol or trehalose synthesis genes, such as trehalose-phosphate synthase or trehalose-phosphate phosphatase (WO 97/42326), or by inhibiting genes such as the trehalase gene (WO 97/50561). Especially preferred nucleic acids are those which encode the transcriptional activator CBF1 from *Arabidopsis thaliana* (GenBank Acc. No.: U77378) or the *Myoxocephalus octodecemspinosus* antifreeze protein (GenBank Acc. No.: AF306348), or functional equivalents of these. For expression in plants, the nucleic acid molecule must be linked operably to a suitable promoter. The plant specific promoter, regulatory element and the terminator of the inventive recombinant expression construct needs not be of plant origin, and may originate from viruses or microorganisms, in particular for example from viruses which attack plant cells.

[0128] An additional subject matter of the invention is the introduction of an inventive intron sequence into a target nucleic acid sequence via homologous recombination (HR). As a prerequisite for the HR between the recombinant expression construct and the genomic target nucleic acid sequence, the recombinant expression construct must contain fragments of the target nucleic acid sequence of sufficient length and homology. In a preferred embodiment of the invention, the intron sequences that has to be inserted into the gene of interest via HR is (within the recombinant expression construct) placed between a pair of DNA sequences identical to the region 5'and 3' to the preferred place of insertion. In this case, the recombinant expression construct can comprises only the intron sequence and the nucleic acid sequences needed to induce the HR event. In a preferred embodiment of the invention, the intron sequence that is flanked by the nucleic acid sequence of the target DNA, contains an expression cassette that enables the expression of an selectable marker protein which allows the selection of transgenic plants in which a homologues or illegitimate recombination had occurred subsequent to the transformation. The expression cassette driving the expression of the selection marker protein can be flanked by HR control sequences that are recognized by specific endonucleases or recombinases, facilitating the removal of the expression cassette from the genome. Such so called marker excision methods e.g. the cre/lox technology permit a tissue-specific, if appropriate inducible, removal of the expression cassette from the genome of the host organism (Sauer B (1998) Methods. 14(4):381-92). In this method, specific flanking sequences (lox sequences), which later allow removal by means of cre recombinase, are attached to the target gene.

[0129] Specifically, the present invention relates to transgenic expression cassettes comprising the following introns with gene expression enhancing properties in plants:

1) The sequence of the first intron (BPSI.1, SEQ ID NO: 1) isolated from the *Oryza sativa* metallothioneine-like gene (Gene Bank accession No. AP002540, Oryza sativa (Japonica cultivar group) genomic DNA, Chromosome 1, PAC clone: P0434B04, gene_id = "P0434B04.31, protein_id ="BAB44010.1", complement joined sequences: 142304.. 142409, 143021..143098, 143683..143747; Hsieh, H.M. et al., RNA expression patterns of a type 2 metallothioneine-like gene from rice. Plant Mol. Biol. 32 (3), 525-529 (1996)). The gene comprises two introns and three exons. The

first intron of the *Oryza sativa* metallothioneine-like gene (BPSI.1, SEQ ID NO:1) is flanked by the 5' (5'-GU-3', base pair (bp) 1-2 in SEQ ID NO:1) and 3' (5'-CAG-3',bp 582-584 in SEQ ID NO:1) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* metallothioneine-like gene (BPSI.1, SEQ ID NO:1) comprises at least 28 bases pairs, more preferably at least 40 bases pairs, most preferably at least 50 base pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively (SEQ ID NO: 82). On nucleotide level, the Oryza sativa metallothionein-like gene shares high homology or identity with the coding region of orthologous genes from other monocotyledonous or dicotyledonous plants e.g. 89% identity to the *Zea mays* CL1155_3 mRNA sequence (acc. No. AY109343), 88% identity to the *Poa secunda* metallothionein-like protein type 2 mRNA (acc. No. AF246982.1), 93% identity to the *Triticum aestivum* metallothioneine mRNA, partial coding sequence (acc. No.AF470355.1), 89% identity to the *Nicotiana plumbaginifolia* metallothionein-like protein mRNA (acc. No. NPU35225), 86% identity to the *Brassica oleracea* cultivar Green King metallothioneine-like protein 2 (acc. No. AF200712), 95% and 88% identity to the *Hordeum vulgare subsp. vulgare* partial mRNA for metallothioneine type2 mt2b (acc. No. HVU511346) and mtb2a (acc. No. HVU511345) genes, respectively (identities have been calculated using the BLASTN 2.2.9 algorithm [May-01-2004] Altschul, Stephen F. et al., (1997), Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25:3389-3402).

2) The sequence of the first intron (BPSI.2, SEQ ID NO:2) isolated from the *Oryza sativa* Sucrose UDP Glucosyltransferase-2 gene (Gene Bank accession No. AC084380, Oryza sativa (Japonica cultivar group) genomic DNA, chromosome 3, BAC OS-JNBa0090P23, gene ID ="OSJNBa0090P23.15",Protein ID=AAK5219.1, complement join (nucleotide 62884 to. 65255, 65350..65594, 65693..66011, 66098..66322, 66427..66593, 66677..66793, 66881..67054, 67136..67231, 67316..67532, 67652..67770, 67896..68088, 68209..68360, 68456..68585, 69314..69453 and 70899..72082). The gene comprises 13 introns and 14 exons. The first intron of the *Oryza sativa* Sucrose UDP Glucosyltransferase-2 gene (BPSI.2, SEQ ID NO: 2) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:2) and 3' (5'-CAG-3',bp 726-728 in SEQ ID NO: 2) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* Sucrose UDP Glucosyltransferase-2 gene (SEQ ID NO:2) comprises at least 19 bases pairs of the sequence 5' to the 5'-splice site and 23 bases pairs of the sequences/exons 3' to the 3'-splice site of the intron (SEQ ID NO: 83). In a particularly preferred embodiment the intron BPSI.2 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively

3) The sequence of the second intron isolated from the *Oryza sativa* Sucrose UDP Glucosyltransferase-2 gene (BPSI.3, SEQ ID NO:3). Said the second intron is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:3) and 3' (5'-CAG-3',bp 93-95 in SEQ ID NO: 3) splice sites.
In a preferred embodiment of the invention, the second intron of the *Oryza sativa* Sucrose UDP Glucosyltransferase-2 gene (SEQ ID NO:3) comprises at least 25 bases pairs of the sequence 5' to the 5'-splice site and 30 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 84). In a particularly preferred embodiment the intron BPSI.3 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively. On nucleotide level, the *Oryza sativa* Sucrose UDP Glucosyltransferase-2 gene shares high homology or identity with the coding region of orthologous genes from other monocotyledonous or dicotyledonous plants e.g. 88% identity to the *Zea mays* sucrose synthase (Sus1) mRNA (acc. No. L22296.1), 85% identity to the *Triticum aestivum* mRNA for sucrose synthase type 2 (acc. No. AJ000153), 85% identity to the *H. vulgare* mRNA for sucrose synthase (acc No. X69931), 80% identity to the *Saccharum officinarum* sucrose synthase-2 mRNA (acc No. AF263384.1), 95% identity to the Rice mRNA for sucrose synthase (S464 gene), partial sequence (acc. No. D10418), 79% identity to the *Glycine max* sucrose synthase mRNA (acc. No. AF03231). Identities have been calculated using the BLASTN 2.2.9 algorithm [May-01-2004] Altschul, Stephen F. et al., (1997), Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25:3389-3402).

4) The sequence of the eighth intron (BPSI.5, SEQ ID NO:5) isolated from the *Oryza sativa* gene encoding for the Sucrose transporter (Gene Bank accession No. AF 280050). Said the eighth intron (SEQ ID NO:5) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:5) and 3' (5'-CAG-3',bp 223-225 in SEQ ID NO: 5) splice sites. In a preferred embodiment of the invention, the eighth intron of the *Oryza sativa* gene encoding for the Sucrose transporter (SEQ ID NO:5) comprises at least 35 bases pairs of the sequence 5' to the 5'-splice site and 30 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 86). In a particularly preferred embodiment the intron BPSI.5comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively. In a more preferred embodiment, the 5' and 3' splice sites of the eighth intron (BPSI.5, SEQ ID NO:5) are modified in order to match the plant consensus sequences for 5' splice sites 5'-AG::GTAAGT-3' (SEQ ID NO: 80) and 3' splice sites 5'-CAG::GT-3' (SEQ ID NO: 81) using a PCR mutagenesis

approach (SEQ ID NO:87).

5) The sequence of the fourth intron (BPSI.6, SEQ ID NO:6) isolated from the *Oryza sativa* gene (Gene Bank accession No. BAA94221) encoding for an unknown protein with homology to the *A. thaliana* chromosome II sequence from clones T22O13, F12K2 encoding for a putative lipase (AC006233). Said the fourth intron (SEQ ID NO: 6) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:6) and 3' (5'-CAG-3', bp 768-770 in SEQ ID NO:6) splice sites. In a preferred embodiment of the invention, the fourth intron of the *Oryza sativa* gene (accession No. BAA94221) (SEQ ID NO:6) comprises at least 34 bases pairs of the sequence 5' to the 5'-splice site and 34 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 88). In a particularly preferred embodiment the intron BPSI.6 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively. In a more preferred embodiment, the 5' and 3' splice sites of fourth intron (BPSI.6, SEQ ID NO:6) are modified in order to match the plant consensus sequences for 5' splice sites 5'-AG::GTAAGT-3' (SEQ ID NO: 80) and 3' splice sites 5'-CAG::GT-3' (SEQ ID NO: 81) using a PCR mutagenesis approach (SEQ ID NO:89).

6) The sequence of the fourth intron (BPSI.7, SEQ ID NO:7) isolated from the *Oryza sativa* gene (accession No. BAB90130) encoding for a putative cinnamyl-alcohol dehydrogenase. Said the fourth intron (SEQ ID NO:7) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:7) and 3' (5'-CAG-3', 713-715 bp in SEQ ID NO: 7) splice sites. In a preferred embodiment of the invention, the fourth intron of the *Oryza sativa* gene (accession No. BAB90130) (SEQ ID NO:7) comprises at least 34 bases pairs of the sequence 5' to the 5'-splice site and 26 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 90). In a particularly preferred embodiment the intron BPSI.7 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively. In a more preferred embodiment, the 5' and 3' splice sites of the fourth intron (BPSI.7, SEQ ID NO:7) are modified in order to match the plant consensus sequences for 5' splice sites 5'-AG::GTAAGT-3' (SEQ ID NO: 80) and 3' splice sites 5'-CAG::GT-3' (SEQ ID NO: 81) using a PCR mutagenesis approach (SEQ ID NO:91).

7) The sequence of the third intron (BPSI.10, SEQ ID NO:10) isolated from the *Oryza sativa* gene (accession No. AP003300) encoding for a putative protein kinase. Said the third intron (SEQ ID NO:10) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:10) and 3' (5'-CAG-3', 536-538 bp in SEQ ID NO: 10) splice sites. In a preferred embodiment of the invention, the third intron of the *Oryza sativa* gene (accession No. AP003300) (SEQ ID NO:10) comprises at least 31 bases pairs of the sequence 5' to the 5'-splice site and 31 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 94). In a particularly preferred embodiment the intron BPSI.10 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively. In a more preferred embodiment, the 5' and 3' splice sites of the third intron (BPSI.10, SEQ ID NO:10) are modified in order to match the plant consensus sequences for 5' splice sites 5'-AG::GTAAGT-3' (SEQ ID NO: 80) and 3' splice sites 5'-CAG::GT-3' (SEQ ID NO: 81) using a PCR mutagenesis approach (SEQ ID NO:95).

8) The sequence of the first intron (BPSI.11, SEQ ID NO:11) isolated from the *Oryza sativa* gene (accession No. L37528) encoding for a MADS3 box protein. Said the first intron (SEQ ID NO:11) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:11) and 3' (5'-CAG-3', bp 329-331 in SEQ ID NO: 11) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. L37528) (SEQ ID NO:11) comprises at least 35 bases pairs of the sequence 5' to the 5'-splice site and 34 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 96). In a particularly preferred embodiment the intron BPSI.11 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively. In a more preferred embodiment, the 5' and 3' splice sites of the first intron (BPSI.11, SEQ ID NO:11) are modified in order to match the plant consensus sequences for 5' splice sites 5'-AG::GTAAGT-3' (SEQ ID NO: 80) and 3' splice sites 5'-CAG::GT-3' (SEQ ID NO: 81) using a PCR mutagenesis approach (SEQ ID NO:97).

9) The sequence of the first intron (BPSI.12, SEQ ID NO:12) isolated from the *Oryza sativa* gene (accession No. CB625805) encoding for a putative Adenosylmethionine decarboxylase. Said the first intron (SEQ ID NO:12) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:12) and 3' (5'-CAG-3', bp 959-961 in SEQ ID NO: 12) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. CB625805) (SEQ ID NO:12) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 26 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 98). In a particularly preferred embodiment the intron BPSI.12 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

10) The sequence of the first intron (BPSI.13, SEQ ID NO:13) isolated from the *Oryza sativa* gene (accession No. CF297669) encoding for an Aspartic proteinase. Said the first intron (SEQ ID NO:13) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:13) and 3' (5'-CAG-3', bp 593-595 in SEQ ID NO: 13) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. CF297669) (SEQ ID NO:13) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 24 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 99). In a particularly preferred embodiment the intron BPSI.13 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

11) The sequence of the first intron (BPSI.14, SEQ ID NO:14) isolated from the *Oryza sativa* gene (accession No. CB674940) encoding for a Lec14b protein. Said the first intron (SEQ ID NO:14) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:14) and 3' (5'-CAG-3', bp 143-145 in SEQ ID NO: 14) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. CB674940) (SEQ ID NO:14) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 25 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 100). In a particularly preferred embodiment the intron BPSI.14 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

12) The sequence of the first intron (BPSI.15, SEQ ID NO:15) isolated from the 5' UTR of the *Oryza sativa* gene (accession No. BAD37295.1) encoding for a putative SalT protein precursor. Said the first intron (SEQ ID NO:15) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:15) and 3' (5'-CAG-3', bp 312-314 in SEQ ID NO: 15) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No.BAD37295.1) (SEQ ID NO:15) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 25 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 101). In a particularly preferred embodiment the intron BPSI.15 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

13) The sequence of the first intron (BPSI.16, SEQ ID NO:16) isolated from the *Oryza sativa* gene (accession No. BX928664) encoding for a putative reticulon. Said the first intron (SEQ ID NO:16) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:16) and 3' (5'-CAG-3', bp 650-652 in SEQ ID NO: 16) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. BX928664) (SEQ ID NO:16) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 23 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 102). In a particularly preferred embodiment the intron BPSI.16 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

14) The sequence of the first intron (BPSI.17, SEQ ID NO:17) isolated from the *Oryza sativa* gene (accession No. AA752970) encoding for a glycolate oxidase. Said the first intron (SEQ ID NO:17) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:17) and 3' (5'-CAG-3', bp 266-268 in SEQ ID NO:17) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. AA752970) (SEQ ID NO:17) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 35 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 103). In a particularly preferred embodiment the intron BPSI.17 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

15) The sequence of the first intron (BPSI.18, SEQ ID NO:18) isolated from the *Oryza sativa* clone GI 40253643 (accession No. AK064428) is similar to AT4g33690. Said the first intron (SEQ ID NO:18) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:18) and 3' (5'-CAG-3', bp 544-546 in SEQ ID NO:18) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. AK064428) (SEQ ID NO:18) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 21 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 104). In a particularly preferred embodiment the intron BPSI.18 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

16) The sequence of the first intron (BPSI.19, SEQ ID NO:19) isolated from the *Oryza sativa* clone GI 51091887 (accession No. AK062197)). Said the first intron (SEQ ID NO:19) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:19) and 3' (5'-CAG-3', bp 810-812 in SEQ ID NO:19) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. AK062197) (SEQ ID NO:19) comprises at least 26 bases

pairs of the sequence 5' to the 5'-splice site and 26 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 105). In a particularly preferred embodiment the intron BPSI.19 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

17) The sequence of the first intron (BPSI.20, SEQ ID NO:20) isolated from the *Oryza sativa* gene (accession No. CF279761) encoding for a hypothetical protein clone (GI 33657147). Said the first intron (SEQ ID NO:20) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:20) and 3' (5'-CAG-3', bp 369-371 in SEQ ID NO:20) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. CF279761) (SEQ ID NO:20) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 27 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 106). In a particularly preferred embodiment the intron BPSI.20 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

18) The sequence of the first intron (BPSI.21, SEQ ID NO:21) isolated from the *Oryza sativa* gene (accession No. CF326058) encoding for a putative membrane transporter. Said the first intron (SEQ ID NO:21) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:21) and 3' (5'-CAG-3', bp 720-722 in SEQ ID NO:21) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. CF326058) (SEQ ID NO:21) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 25 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 107). In a particularly preferred embodiment the intron BPSI.21 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

19) The sequence of the first intron (BPSI.22, SEQ ID NO:22) isolated from the *Oryza sativa* gene (accession No. C26044) encoding for a putative ACT domain repeat protein. Said the first intron (SEQ ID NO:22) is flanked by the 5' (5'-GU-3', bp 1-2 in SEQ ID NO:22) and 3' (5'-CAG-3', bp 386-388 in SEQ ID NO:22) splice sites. In a preferred embodiment of the invention, the first intron of the *Oryza sativa* gene (accession No. C26044) (SEQ ID NO:22) comprises at least 26 bases pairs of the sequence 5' to the 5'-splice site and 28 bases pairs of the sequences 3' to the 3'-splice site of the intron (SEQ ID NO: 108). In a particularly preferred embodiment the intron BPSI.22 comprises at least 40 bases pairs, more preferably at least 50 bases pairs of the sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron, respectively.

Table 1: Genes from which the introns of the invention are preferably isolated, putative function of said genes, cDNA and the protein encoded by said genes.

Intron	Rice GI number	Accesion No.	SEQ ID NO.	Sequence homology
BPSI.1		AP002540	1	metallothioneine-like gene
BPSI.2		AC084380	2	Sucrose UDP Glucosyltransferase-2 gene, first Intron
BPSI.3		AC084380	3	Sucrose UDP Glucosyltransferase-2 gene, second Intron
BPSI.4		AC084380	4	Sucrose UDP Glucosyltransferase-2 gene, third Intron
BPSI.5	9624451	AF280050	5	Sucrose transporter

(continued)

Intron	Rice GI number	Accesion No.	SEQ ID NO.	Sequence homology
BPSI.6	7523493	BAA94221	6	Similar to A. thaliana chromosome II sequence from clones T22O13, F12K2; putative lipase (AC006233)
BPSI.7	20161203	BAB90130	7	putative cinnamyl-alcohol dehydrogenase
BPSI.10	20160990	AP003300	10	Putative protein kinase
BPSI.11	886404	L37528	11	MADS3 box protein
BPSI.12	29620794	CB625805	12	putative Adenosylmethionine decarboxylase
BPSI.13	33666702	CF297669	13	Aspartic proteinase
BPSI.14	29678665	CB674940	14	Lec14b protein
BPSI.15	51535011	BAD37295	15	putative SalT protein precursor
BPSI.16	41883853	BX928664	16	Putative Reticulon
BPSI.17	2799981	AA752970	17	Glycolate oxidase
BPSI.18	40253643	AK06442	18	Putative non-coding (Similar to AT4g33690)
BPSI.19	51091887	AK062197	19	Putative non-coding
BPSI.20	33657147	CF279761	20	Hypothetical protein
BPSI.21	33800379	CF326058	21	Putative membrane transporter
BPSI.22	2309889	C26044	22	Putative ACT domain repeat protein

[0130] It is disclosed by the examples of this invention, that the inventive introns with the SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10 and 11 have an impact on the expression rate of the GUS gene in transient expression assays and stable transformed plants, respectively. It could be shown that the inclusion of said Introns into the 5' UTR of the GUS gene has led to a strong enhancement in the expression rate of this gene in transiently and stable transformed cell, respectively, compared to a control construct that lacks the first intron (see examples 1.6.1 (table 7), 1.6.2 (table 8), 2.4 (table 15). The expression enhancing properties of the introns with the SEQ ID NOs: 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 can be demonstrated by performing the above described transient or stable expression assays.

[0131] Functional equivalents of the inventive introns can be identified via homology searches in nucleic acid databases or via DNA hybridization (screening of genomic DNA libraries) using a fragment of at least 50 consecutive base pairs of the nucleic acid molecule described by any of the SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 and stringent hybridization conditions. In a preferred embodiment of the present invention the stringent hybridizing conditions can be chosen as follows: The hybridization puffer contains Formamide, NaCl and PEG 6000 (Polyethyleneglykol MW 6000). Formamide has a destabilizing effect on double strand nucleic acid molecules, thereby, when used in hybridization buffer, allowing the reduction of the hybridization temperature to 42°C without reducing the hybridization stringency. NaCl has a positive impact on the renaturation-rate of a DNA duplex and the hybridization efficiency of a DNA probe with its complementary DNA target. PEG increases the viscosity of the hybridization buffer,

which has in principle a negative impact on the hybridization efficiency. The composition of the hybridization buffer is as follows:

250 mM Sodium phosphate-buffer pH 7,2
1 mM EDTA (ethylenediaminetetraacetic acid)
7 % SDS (g/v) (sodium dodecyl sulfate)
250 mM NaCl (Sodiumchloride) 10 μg/ml single stranded DNA 5 % Polyethylenglykol (PEG) 6000 40 % Formamide

[0132] The hybridization is preferably performed over night at 42°C. In the morning, the hybridized filter will be washed 3 x for 10 minutes with 2xSSC + 0,1 % SDS. Hybridization should advantageously be carried out with fragments of at least 50, 60, 70 or 80 bp, preferably at least 90 bp. In an especially preferred embodiment, the hybridization should be carried out with the entire nucleic acid sequence with conditions described above.

[0133] Combination of the introns of the invention with different plant promoters has clearly demonstrated their expression enhancing and/or modulating properties. In a preferred embodiment of the invention the recombinant DNA expression construct comprises (functionally linked to an intron of the invention) a promoter sequence functioning in plants or plant cells selected from the group consisting of

a) the rice chloroplast protein 12 (Os.CP12) promoter as described by nucleotide 1 to 854 of SEQ ID NO: 113 (the "fragment"), or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined in the paragraph above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of said fragment, and

b) the maize hydroxyproline-rich glycoprotein (Zm.HRGP) promoter as described by nucleotide 1 to 1184 of SEQ ID NO: 114, or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined in the paragraph above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of said fragment, and

c) the rice p-caffeoyl-CoA 3-O-methyltransferase (Os.CCoAMT1) promoter as described by nucleotide 1 to 1034 of SEQ ID NO: 115, or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined in the paragraph above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of said fragment, and

d) the maize Globulin-1 (Zm.Glb1) promoter (W64A) as described by nucleotide 1 to 1440 of SEQ ID NO: 116, or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined in the paragraph above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of said fragment, and

e) the putative Rice H+-transporting ATP synthase (Os.V-ATPase) promoter as described by nucleotide 1 to 1589 of SEQ ID NO: 117, or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined in the paragraph above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of said fragment, and

f) the putative rice C-8,7 sterol isomerase (Os.C8,7 SI) promoter as described by nucleotide 1 to 796 of SEQ ID NO: 118, or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined in the paragraph above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most

preferably at least 400 or 500) consecutive nucleotides of said fragment, and
g) the maize lactate dehydrogenase (Zm.LDH) promoter as described by nucleotide 1 to 1062 of SEQ ID NO: 119, or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined in the paragraph above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of said fragment, and
h) the rice Late Embryogenesis Abundant (Os.Lea) promoter as described by nucleotide 1 to 1386 of SEQ ID NO: 121, or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined in the paragraph above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of said fragment.

[0134] Preferably said expression construct is comprising a combination of one of the above defined promoters with at least one intron selected from the group consisting of

i) the BPSI.1 intron as described by nucleotide 888 to 1470 of SEQ ID NO: 113 or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of said fragment and
ii) the BPSI.5 intron as described by nucleotide 1068 to 1318 of SEQ ID NO: 120, or a sequence having at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) identity to said fragment, or a sequence hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined above) to said fragment, or a sequence comprising at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of said fragment.

[0135] More preferably expression construct is comprising a combination of promoter and intron selected from the group consisting of

i) sequences as described by any of SEQ ID NO: 113, 114, 115, 116, 117, 118, 119, 120, or 121, and
ii) sequences having at least 50 (preferably at least 100, more preferably at least 200 or 300, most preferably at least 400 or 500) consecutive nucleotides of a sequence described by any of SEQ ID NOs: 113, 114, 115, 116, 117, 118, 119, 120, or 121, and
iii) sequences having an identity of at least 60% (preferably at least 70% or 80%, more preferably at least 90% or 95%, most preferably at least 98% or 99%) to a sequence described by any of SEQ ID NOs: 113, 114, 115, 116, 117, 118, 119, 120, or 121, and
iv) sequences hybridizing under stringent conditions (preferably under conditions equivalent to the high stringency conditions defined above) with sequence described by any of SEQ ID NOs: 113, 114, 115, 116, 117, 118, 119, 120, or 121.

[0136] A preferred subject matter of the invention, is a vector, preferably a plant transformation vector, containing an inventive recombinant expression construct. The expression cassette can be introduced into the vector via a suitable restriction cleavage site. The plasmid formed is first introduced into *E.coli.* Correctly transformed *E.coli* are selected, grown, and the recombinant plasmid is obtained by the methods familiar to the skilled worker. Restriction analysis and sequencing may serve to verify the cloning step. Preferred vectors are those, which make possible stable integration of the expression cassette into the host genome. An expression construct according to the invention can advantageously be introduced into cells, preferably into plant cells, using vectors. In one embodiment, the methods of the invention involve transformation of organism or cells (e.g. plants or plant cells) with a transgenic expression vector comprising at least a transgenic expression cassette of the invention. The methods of the invention are not limited to the expression vectors disclosed herein. Any expression vector which is capable of introducing a nucleic acid sequence of interest into a plant cell is contemplated to be within the scope of this invention. Typically, expression vectors comprise the transgenic expression cassette of the invention in combination with elements which allow cloning of the vector into a bacterial or phage host. The vector preferably, though not necessarily, contains an origin of replication which is functional in a broad range of prokaryotic hosts. A selectable marker is generally, but not necessarily, included to allow selection of cells

bearing the desired vector. Preferred are those vectors that allowing a stable integration of the expression construct into the host genome. In the case of injection or electroporation of DNA into plant cells, the plasmid used need not meet any particular requirements. Simple plasmids such as those of the pUC series can be used. If intact plants are to be regenerated from the transformed cells, it is necessary for an additional selectable marker gene to be present on the plasmid. A variety of possible plasmid vectors are available for the introduction of foreign genes into plants, and these plasmid vectors contain, as a rule, a replication origin for multiplication in *E.coli* and a marker gene for the selection of transformed bacteria. Examples are pBR322, pUC series, M13mp series, pACYC184 and the like. The expression construct can be introduced into the vector via a suitable restriction cleavage site. The plasmid formed is first introduced into *E.coli.* Correctly transformed *E.coli* are selected and grown, and the recombinant plasmid is obtained by methods known to the skilled worker. Restriction analysis and sequencing can be used for verifying the cloning step.

[0137] Depending on the method by which DNA is introduced, further genes may be necessary on the vector plasmid.

[0138] *Agrobacterium tumefaciens* and *A. rhizogenes* are plant-pathogenic soil bacteria, which genetically transform plant cells. The Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes,* respectively, carry genes responsible for genetic transformation of the plant (Kado (1991) Crit Rev Plant Sci 10:1). Vectors of the invention may be based on the *Agrobacterium* Ti- or Ri-plasmid and may thereby utilize a natural system of DNA transfer into the plant genome. As part of this highly developed parasitism *Agrobacterium* transfers a defined part of its genomic information (the T-DNA; flanked by about 25 bp repeats, named left and right border) into the chromosomal DNA of the plant cell (Zupan (2000) Plant J 23(1):11-28). By combined action of the so-called *vir* genes (part of the original Ti-plasmids) said DNA-transfer is mediated. For utilization of this natural system, Ti-plasmids were developed which lack the original tumor inducing genes ("disarmed vectors"). In a further improvement, the so called "binary vector systems", the T-DNA was physically separated from the other functional elements of the Ti-plasmid (e.g., the *vir* genes), by being incorporated into a shuttle vector, which allowed easier handling (EP-A 120 516; US 4.940.838). These binary vectors comprise (beside the disarmed T-DNA with its border sequences), prokaryotic sequences for replication both in *Agrobacterium* and *E. coli.* It is an advantage of *Agrobacterium-mediated* transformation that in general only the DNA flanked by the borders is transferred into the genome and that preferentially only one copy is inserted. Descriptions of *Agrobacterium* vector systems and methods for *Agrobacterium-mediated* gene transfer are known in the art (Miki 1993, "Procedures for Introducing Foreign DNA into Plants" in METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY; pp.67-88; Gruber 1993, "Vectors for Plant Transformation," in METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY; pp.89-119; Moloney (1989) Plant Cell Reports 8: 238-242). The use of T-DNA for the transformation of plant cells has been studied and described intensively (EP 120516; Hoekema 1985, In: The Binary Plant Vector System, Offsetdrukkerij Kanters B.V., Alblasserdam, Chapter V; Fraley (1985) CRC Crit. Rev. Plant. Sci. 4:1-45; and An (1985) EMBO J. 4:277-287). Various binary vectors are known, some of which are commercially available such as, for example, pBIN19 (Clontech Laboratories, Inc. U.S.A.). Hence, for *Agrobacterium-mediated* transformation the transgenic expression construct of the invention is integrated into specific plasmids, either into a shuttle or intermediate vector, or into a binary vector. If a Ti or Ri plasmid is to be used for the transformation, at least the right border, but in most cases the right and left border, of the Ti or Ri plasmid T-DNA is linked to the transgenic expression construct to be introduced in the form of a flanking region. Binary vectors are preferably used. Binary vectors are capable of replication both in *E.coli* and in *Agrobacterium.* They may comprise a selection marker gene and a linker or polylinker (for insertion of e.g. the expression construct to be transferred) flanked by the right and left T-DNA border sequence. They can be transferred directly into *Agrobacterium* (Holsters (1978) Mol Gen Genet 163:181-187). The selection marker gene permits the selection of transformed agrobacteria and is, for example, the *npt*II gene, which confers resistance to kanamycin. The *Agrobacterium* which acts as host organism in this case should already contain a plasmid with the *vir* region. The latter is required for transferring the T-DNA to the plant cell. An *Agrobacterium* transformed in this way can be used for transforming plant cells. The use of T-DNA for transforming plant cells has been studied and described intensively (EP 120 516; Hoekema (1985) Nature 303:179-181; An (1985) EMBO J. 4:277-287; see also below). Common binary vectors are based on "broad host range"-plasmids like pRK252 (Bevan (1984) Nucl Acid Res 12:8711-8720) or pTJS75 (Watson (1985) EMBO J 4(2):277-284) derived from the P-type plasmid RK2. Most of these vectors are derivatives of pBIN19 (Bevan 1984, Nucl Acid Res 12:8711-8720). Various binary vectors are known, some of which are commercially available such as, for example, pB1101.2 or pBIN19 (Clontech Laboratories, Inc. USA). Additional vectors were improved with regard to size and handling (e.g. pPZP; Hajdukiewicz (1994) Plant Mol Biol 25:989-994). Improved vector systems are described also in WO 02/00900. In a preferred embodiment, *Agrobacterium* strains for use in the practice of the invention include octopine strains, e.g., LBA4404 or agropine strains, e.g., EHA101 or EHA105. Suitable strains of *A. tumefaciens* for DNA transfer are for example EHA101 pEHA101 (Hood (1986) J Bacteriol 168:1291-1301), EHA105[pEHA105] (Li (1992) Plant Mol Biol 20: 1037-1048), LBA4404[pAL4404] (Hoekema (1983) Nature 303:179-181), C58C1[pMP90] (Koncz (1986) Mol Gen Genet 204:383-396), and C58C1[pGV2260] (Deblaere (1985) Nucl Acids Res 13:4777-4788. Other suitable strains are *Agrobacterium tumefaciens* C58, a nopaline strain. Other suitable strains are *A. tumefaciens* C58C1 (Van Larebeke (1974) Nature 252:169-170, A136 (Watson (1975) J. Bacteriol 123:255-264) or LBA4011 (Klapwijk (1980) J. Bacteriol. 141: 128-136 In a preferred embodiment, the *Agrobacterium* strain used to transform the plant tissue pre-cultured with the

plant phenolic compound contains a L,L-succinamopine type Ti-plasmid, preferably disarmed, such as pEHA101. In another preferred embodiment, the *Agrobacterium* strain used to transform the plant tissue pre-cultured with the plant phenolic compound contains an octopine-type Ti-plasmid, preferably disarmed, such as pAL4404. Generally, when using octopine-type Ti-plasmids or helper plasmids, it is preferred that the *virF* gene be deleted or inactivated (Jarchow (1991) Proc. Natl. Acad. Sci. USA 88:10426-10430). In a preferred embodiment, the *Agrobacterium* strain used to transform the plant tissue pre-cultured with the plant phenolic compound such as acetosyringone. The method of the invention can also be used in combination with particular *Agrobacterium* strains, to further increase the transformation efficiency, such as *Agrobacterium* strains wherein the *vir* gene expression and/or induction thereof is altered due to the presence of mutant or chimeric *vir*A or *vir*G genes (e.g. Hansen (1994) Proc. Natl. Acad. Sci. USA 91:7603-7607;Chen 1991 J. Bacteriol. 173:1139-1144; Scheeren-Groot (1994) J. Bacteriol 176:6418-6426). A binary vector or any other vector can be modified by common DNA recombination techniques, multiplied in *E. coli,* and introduced into *Agrobacterium* by e.g., electroporation or other transformation techniques (Mozo (1991) Plant Mol. Biol. 16:917-918). *Agrobacterium* is grown and used as described in the art. The vector comprising *Agrobacterium* strain may, for example, be grown for 3 days on YP medium (5 g/L yeast extract, 10 g/L peptone, 5 g/L Nail, 15 g/L agar, pH 6.8) supplemented with the appropriate antibiotic (e.g., 50 mg/L spectinomycin). Bacteria are collected with a loop from the solid medium and resuspended. An additional subject matter of the invention relates to transgenic non-human organisms transformed with at least one vector containing a transgenic expression construct of the invention. In a preferred embodiment the invention relates to bacteria, fungi, yeasts, more preferably to plants or plant cell. In a preferred embodiment of the invention, the transgenic organism is a monocotyledonous plant. In a yet more preferred embodiment, the monocotyledonous plant is selected from the group consisting of the genera *Hordeum, Avena, Secale, Triticum, Sorghum, Zea, Saccharum* and *Oryza,* very especially preferred are plants selected from the group consisting of *Hordeum vulgare, Triticum aestivum, Triticum aestivum subsp.spelta, Triticale, Avena sativa, Secale cereale, Sorghum bicolor, Saccharum officinarum, Zea mays* and *Oryza sativa* transformed with the inventive vectors or containing the inventive recombinant expression constructs. Preferred bacteria are bacteria of the genus *Escherichia, Erwinia, Agrobacterium, Flavobacterium, Alcaligenes or cyanobacteria,* for example of the genus *Synechocystis.* Especially preferred are microorganisms which are capable of infecting plants and thus of transferring the constructs according to the invention. Preferred microorganisms are those from the genus *Agrobacterium* and, in particular, the species *Agrobacterium tumefaciens.* Preferred yeasts are *Candida, Saccharomyces, Hansenula or Pichia.* Preferred fungi are *Aspergillus, Trichoderma, Ashbya, Neurospora, Fusarium, Beauveria* or other fungi. Plant organisms are furthermore, for the purposes of the invention, other organisms which are capable of photosynthetic activity such as, for example, algae or cyanobacteria, and also mosses. Preferred algae are green algae such as, for example, algae of the genus *Haematococcus, Phaedactylum tricornatum, Volvox* or *Dunaliella.* Furthermore the invention relates cell cultures, tissues, organs (e.g., leaves, roots and the like in the case of plant organisms), or propagation material derived from transgenic non-human organisms like bacteria, fungi, yeasts, plants or plant cells transformed with at least one vector containing a transgenic expression construct of the invention.

[0139] An additional subject matter of the invention relates to a method for providing an expression cassette for enhanced expression of a nucleic acid in a plant or a plant cell, comprising the step of functionally linking the inventive introns to a plant expression cassette not comprising said intron. In a yet another preferred embodiment, the invention relates to a method for enhancing the expression of a nucleic acid sequence in a plant or a plant cell, comprising functionally linking the inventive introns to said nucleic acid sequence. Preferably, the method for providing an expression cassette for enhanced expression of a nucleic acid in a plant or a plant cell and the method for enhancing the expression of a nucleic acid sequence in a plant or a plant cell further comprises the steps of

i) providing an recombinant expression cassette, wherein the nucleic acid sequence is functionally linked with a promoter sequence functional in plants and with an intron sequence selected from the group consisting of SEQ ID NOs: 1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22,
ii) introducing said recombinant expression into a plant cell or a plant,
iii) identifying or selecting the transgenic plant cell comprising said transgenic expression construct. In another preferred embodiment, the above-described method further comprises the steps of
iv) regenerating transgenic plant tissue from the transgenic plant cell. In an alternative preferred embodiment, the method further comprises
v) regenerating a transgenic plant from the transgenic plant cell.

[0140] The generation of a transformed organism or a transformed cell requires introducing the DNA in question into the host cell in question. A multiplicity of methods is available for this procedure, which is termed transformation (see also Keown (1990) Methods in Enzymology 185:527-537). For example, the DNA can be introduced directly by microinjection or by bombardment via DNA-coated microparticles. Also, the cell can be permeabilized chemically, for example using polyethylene glycol, so that the DNA can enter the cell by diffusion. The DNA can also be introduced by protoplast fusion with other DNA-containing units such as minicells, cells, lysosomes or liposomes. Another suitable method of

introducing DNA is electroporation, where the cells are permeabilized reversibly by an electrical pulse. Methods for introduction of a transgenic expression construct or vector into plant tissue may include but are not limited to, e.g., electroinjection (Nan (1995) In "Biotechnology in Agriculture and Forestry," Ed. Y. P. S. Bajaj, Springer-Verlag Berlin Heidelberg 34:145-155; Griesbach (1992) Hort Science 27:620); fusion with liposomes, lysosomes, cells, minicells or other fusible lipid-surfaced bodies (Fraley (1982) Proc. Natl. Acad. Sci. USA 79:1859-1863); polyethylene glycol (Krens (1982) Nature 296:72-74); chemicals that increase free DNA uptake; transformation using virus, and the like. Furthermore, the biolistic method with the gene gun, electroporation, incubation of dry embryos in DNA-containing solution, and microinjection may be employed. Protoplast based methods can be employed (e.g., for rice), where DNA is delivered to the protoplasts through liposomes, PEG, or electroporation (Shimamoto (1989) Nature 338:274-276; Datta (1990) Bio/Technology 8:736-740). Transformation by electroporation involves the application of short, high-voltage electric fields to create "pores" in the cell membrane through which DNA is taken-up. These methods are - for example - used to produce stably transformed monocotyledonous plants (Paszkowski (1984) EMBO J 3:2717-2722; Shillito (1985) Bio/Technology, 3:1099-1103; Fromm (1986) Nature 319:791-793) especially from rice (Shimamoto (1989) Nature 338: 274-276; Datta (1990) Bio/Technology 8:736-740; Hayakawa (1992) Proc Natl Acad Sci USA 89:9865-9869). Particle bombardment or "biolistics" is a widely used method for the transformation of plants, especially monocotyledonous plants. In the "biolistics" (microprojectile-mediated DNA delivery) method microprojectile particles are coated with DNA and accelerated by a mechanical device to a speed high enough to penetrate the plant cell wall and nucleus (WO 91/02071). The foreign DNA gets incorporated into the host DNA and results in a transformed cell. There are many variations on the "biolistics" method (Sanford (1990) Physiologia Plantarium 79:206-209; Fromm (1990) Bio/Technology 8:833-839; Christou (1988) Plant Physiol 87:671-674; Sautter (1991) Bio/Technology 9:1080-1085). The method has been used to produce stably transformed monocotyledonous plants including rice, maize, wheat, barley, and oats (Christou (1991) Bio/Technology 9:957-962; Gordon-Kamm (1990) Plant Cell 2:603-618; Vasil (1992) Bio/Technology 10: 667-674, (1993) Bio/Technology 11:1153-1158; Wan (1994) Plant Physiol. 104:3748; Somers (1992) Bio/Technology 10:1589-1594). In addition to these "direct" transformation techniques, transformation can also be effected by bacterial infection by means of *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes.* These strains contain a plasmid (Ti or Ri plasmid) which is transferred to the plant following *Agrobacterium* infection. Part of this plasmid, termed T-DNA (transferred DNA), is integrated into the genome of the plant cell (see above for description of vectors). To transfer the DNA to the plant cell, plant explants are cocultured with a transgenic *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes.* Starting from infected plant material (for example leaf, root or stem sections, but also protoplasts or suspensions of plant cells), intact plants can be generated using a suitable medium which may contain, for example, antibiotics or biocides for selecting transformed cells. The plants obtained can then be screened for the presence of the DNA introduced, in this case the expression construct according to the invention. As soon as the DNA has integrated into the host genome, the genotype in question is, as a rule, stable and the insertion in question is also found in the subsequent generations. The plants obtained can be cultured and hybridized in the customary fashion. Two or more generations should be grown in order to ensure that the genomic integration is stable and hereditary. The abovementioned methods are described (for example, in Jenes (1983) Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, edited by Kung & Wu, Academic Press 128-143; and in Potrykus (1991) Ann Rev Plant Physiol Plant Mol Biol 42:205-225). One of skill in the art knows that the efficiency of transformation by *Agrobacterium* may be enhanced by using a number of methods known in the art. For example, the inclusion of a natural wound response molecule such as acetosyringone (AS) to the *Agrobacterium* culture has been shown to enhance transformation efficiency with *Agrobacterium tumefaciens* (Shahla (1987) Plant Mol. Biol. 8:291-298). Alternatively, transformation efficiency may be enhanced by wounding the target tissue to be transformed. Wounding of plant tissue may be achieved, for example, by punching, maceration, bombardment with microprojectiles, etc. (see, e.g., Bidney (1992) Plant Molec. Biol. 18: 301-313). A number of other methods have been reported for the transformation of plants (especially monocotyledonous plants) including, for example, the "pollen tube method" (WO 93/18168; Luo (1988) Plant Mol. Biol. Rep. 6:165-174), macro-injection of DNA into floral tillers (Du (1989) Genet Manip Plants 5:8-12), injection of *Agrobacterium* into developing caryopses (WO 00/63398), and tissue incubation of seeds in DNA solutions (Töpfer (1989) Plant Cell 1:133-139). Direct injection of exogenous DNA into the fertilized plant ovule at the onset of embryogenesis was disclosed in WO 94/00583. WO 97/48814 disclosed a process for producing stably transformed fertile wheat and a system of transforming wheat via *Agrobacterium* based on freshly isolated or pre-cultured immature embryos, embryogenic callus and suspension cells.

[0141] As a rule, the expression construct integrated contains a selection marker, which imparts a resistance to a biocide (for example a herbicide) or an antibiotic such as kanamycin, G 418, bleomycin, hygromycin or phosphinotricin and the like to the transformed plant. The selection marker permits the selection of transformed cells from untransformed cells (McCormick 1986) Plant Cell Reports 5:81-84). The plants obtained can be cultured and hybridized in the customary fashion. Two or more generations should be grown in order to ensure that the genomic integration is stable and hereditary. The abovementioned methods are described (for example, in Jenes 1983; and in Potrykus 1991). As soon as a transformed plant cell has been generated, an intact plant can be obtained using methods known to the skilled worker. Accordingly, the present invention provides transgenic plants. The transgenic plants of the invention are not limited to

plants in which each and every cell expresses the nucleic acid sequence of interest under the control of the promoter sequences provided herein. Included within the scope of this invention is any plant which contains at least one cell which expresses the nucleic acid sequence of interest (e.g., chimeric plants). It is preferred, though not necessary, that the transgenic plant comprises the nucleic acid sequence of interest in more than one cell, and more preferably in one or more tissue. Once transgenic plant tissue, which contains an expression vector, has been obtained, transgenic plants may be regenerated from this transgenic plant tissue using methods known in the art. Species from the following examples of genera of plants may be regenerated from transformed protoplasts: *Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Ciohorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Pisum, Lolium, Zea, Triticum, Sorghum,* and *Datura.* For regeneration of transgenic plants from transgenic protoplasts, a suspension of transformed protoplasts or a Petri plate containing transformed explants is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, somatic embryo formation can be induced in the callus tissue. These somatic embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and plant hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. These three variables may be empirically controlled to result in reproducible regeneration. Plants may also be regenerated from cultured cells or tissues. Dicotyledonous plants which have been shown capable of regeneration from transformed individual cells to obtain transgenic whole plants include, for example, apple (*Malus pumila*), blackberry (*Rubus*), Blackberry/raspberry hybrid (*Rubus*), red raspberry (*Rubus*), carrot (*Daucus carota*), cauliflower (*Brassica oleracea*), celery (*Apium graveolens*), cucumber (*Cucumis sativus*), eggplant (*Solanum melongena*), lettuce (*Lactuca sativa*), potato (*Solanum tuberosum*), rape (*Brassica napus*), wild soybean (*Glycine canescens*), soybean (*Glycine max*), strawberry (*Fragaria ananassa*), tomato (*Lycopersicon esculentum*), walnut (*Juglans regia*), melon (*Cucumis melo*), grape (*Vitis vinifera*), and mango (*Mangifera indica*). Monocotyledonous plants which have been shown capable of regeneration from transformed individual cells to obtain transgenic whole plants include, for example, rice (*Oryza sativa*), rye (*Secale cereale*), and maize (*Zea mays*).

[0142] In addition, regeneration of whole plants from cells (not necessarily transformed) has also been observed in: apricot (*Prunus armeniaca*), asparagus (*Asparagus officinalis*), banana (hybrid *Musa*), bean (*Phaseolus vulgaris*), cherry (hybrid *Prunus*), grape (*Vitis vinifera*), mango (*Mangifera indica*), melon (*Cucumis melo*), ochra (*Abelmoschus esculentus*), onion (hybrid *Allium*), orange (*Citrus sinensis*), papaya (*Carrica papaya*), peach (*Prunus persica*), plum (*Prunus domestica*), pear (*Pyrus communis*), pineapple (*Ananas comosus*), watermelon (*Citrullus vulgaris*), and wheat (*Triticum aestivum*). The regenerated plants are transferred to standard soil conditions and cultivated in a conventional manner. After the expression vector is stably incorporated into regenerated transgenic plants, it can be transferred to other plants by vegetative propagation or by sexual crossing. For example, in vegetatively propagated crops, the mature transgenic plants are propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. In seed propagated crops, the mature transgenic plants are self crossed to produce a homozygous inbred plant which is capable of passing the transgene to its progeny by Mendelian inheritance. The inbred plant produces seed containing the nucleic acid sequence of interest. These seeds can be grown to produce plants that would produce the selected phenotype. The inbred plants can also be used to develop new hybrids by crossing the inbred plant with another inbred plant to produce a hybrid.

[0143] Confirmation of the transgenic nature of the cells, tissues, and plants may be performed by PCR analysis, antibiotic or herbicide resistance, enzymatic analysis and/or Southern blots to verify transformation. Progeny of the regenerated plants may be obtained and analyzed to verify whether the transgenes are heritable. Heritability of the transgene is further confirmation of the stable transformation of the transgene in the plant. The resulting plants can be bred in the customary fashion. Two or more generations should be grown in order to ensure that the genomic integration is stable and hereditary. Corresponding methods are described, (Jenes 1993; Potrykus 1991).

[0144] Also in accordance with the invention are cells, cell cultures, tissues, parts, organs,- such as, for example, roots, leaves and the like in the case of transgenic plant organisms - derived from the above-described transgenic organisms, and transgenic propagation material such as seeds or fruits.

[0145] Preferably, the method for enhancing the expression of a nucleic acid sequence in a plant or a plant cell further comprises,

linking the introns with expression enhancing properties to the expression cassette by insertion via homologous recombination comprising the following steps:

a) providing *in vivo* or *in vitro* a DNA construct comprising said introns flanked by sequences allowing homologous recombination into a pre-existing expression cassette between the promoter and the nucleic acid of said expression cassette,

b) transforming a recipient plant cell comprising said cassette,
regenerating a transgenic plant where said intron has been inserted into the genomic DNA of said promoter nucleic acid construct via homologous recombination.

[0146] Two different ways for the integration of DNA molecules into genomes are possible: Either regions of sequence identity between the partners are used (homologous recombination (HR), "gene targeting") or no sequence-specific requirements have to be fulfilled (illegitimate recombination also referred to as non-homologous end joining (NHEJ)). Gene targeting (GT) is the generation of specific mutations in a genome by homologous recombination-mediated integration of foreign DNA sequences. In contrast to natural recombination processes, one of the recombination partners is artificial and introduced by transformation in gene targeting. The integration of transformed DNA follows pre-existing recombination pathways. Homologous recombination is a reaction between any pair of DNA sequences having a similar sequence of nucleotides, where the two sequences interact (recombine) to form a new recombinant DNA species. The frequency of homologous recombination increases as the length of the shared nucleotide DNA sequences increases, and is higher with linearized plasmid molecules than with circularized plasmid molecules. Homologous recombination can occur between two DNA sequences that are less than identical, but the recombination frequency declines as the divergence between the two sequences increases. Introduced DNA sequences can be targeted via homologous recombination by linking a DNA molecule of interest to sequences sharing homology with endogenous sequences of the host cell. Once the DNA enters the cell, the two homologous sequences can interact to insert the introduced DNA at the site where the homologous genomic DNA sequences were located. Therefore, the choice of homologous sequences contained on the introduced DNA will determine the site where the introduced DNA is integrated via homologous recombination. For example, if the DNA sequence of interest is linked to DNA sequences sharing homology to a single copy gene of a host plant cell, the DNA sequence of interest will be inserted via homologous recombination at only that single specific site. However, if the DNA sequence of interest is linked to DNA sequences sharing homology to a multicopy gene of the host eucaryotic cell, then the DNA sequence of interest can be inserted via homologous recombination at each of the specific sites where a copy of the gene is located. For example, if one wishes to insert a foreign gene into the genomic site where a selected gene is located, the introduced DNA should contain sequences homologous to the selected gene. A double recombination event can be achieved by flanking each end of the DNA sequence of interest (the sequence intended to be inserted into the genome) with DNA sequences homologous to the selected gene. A homologous recombination event involving each of the homologous flanking regions will result in the insertion of the foreign DNA. Thus only those DNA sequences located between the two regions sharing genomic homology become integrated into the genome.

[0147] In the case of gene targeting via homologous recombination, the inventive intron that has to be introduced in the chromosome, preferably in the 5'UTR of a gene (a pre-existing expression cassette), is (for example) located on a DNA construct and is 5' and 3' flanked by nucleic acid sequences of sufficient homology to the target DNA (such an construct is called "gene targeting substrate") in which the intron should be integrated. Said flanking regions must be sufficient in length for making possible recombination. They are, as a rule, in the range of several hundred bases to several kilo bases in length (Thomas KR and Capecchi MR (1987) Cell 51:503; Strepp et al. (1998) Proc Natl Acad Sci USA 95(8):4368-4373). In a preferred embodiment of the invention, the gene targeting substrate comprises an selection marker that is co-integrated with the intron into the genomic region of interest, allowing the selection of recombination events. Preferably, the gene targeting substrate is integrated via a double cross over event between pairs of homologous DNA sequences of sufficient length and homology resulting in the insertion of the intron sequence (and if desired additional nucleic acid sequences e.g. selection marker). Using homologous recombination, a intron of the invention can be placed in the 5' non coding region of the target gene (e.g., an endogenous plant gene) to be transgenically expressed, by linking said intron to DNA sequences which are homologous to, for example, endogenous sequences upstream and/or downstream of the reading frame of the target gene. After a cell has been transformed with the DNA construct in question, the homologous sequences can interact and thus place the intron of the invention at the desired site so that the intron sequence of the invention becomes operably linked to the target gene and constitutes an expression construct of the invention. For homologous recombination or gene targeting, the host organism - for example a plant - is transformed with the recombination construct using the methods described herein, and clones, which have successfully undergone recombination, are selected, for example using a resistance to antibiotics or herbicides. If desirable to target the nucleic acid sequence of interest to a particular locus on the plant genome, site-directed integration of the nucleic acid sequence of interest into the plant cell genome may be achieved by, for example, homologous recombination using Agrobacterium-derived sequences. Generally, plant cells are incubated with a strain of *Agrobacterium* which contains a targeting vector in which sequences that are homologous to a DNA sequence inside the target locus are flanked by *Agrobacterium* transfer-DNA (T-DNA) sequences, as previously described (US 5,501,967, the entire contents of which are herein incorporated by reference).

[0148] One of skill in the art knows that homologous recombination may be achieved using targeting vectors which contain sequences that are homologous to any part of the targeted plant gene, whether belonging to the regulatory

elements of the gene, or the coding regions of the gene. Homologous recombination may be achieved at any region of a plant gene so long as the nucleic acid sequence of regions flanking the site to be targeted is known. Gene targeting is a relatively rare event in higher eucaryotes, especially in plants. Random integrations into the host genome predominate. One possibility of eliminating the randomly integrated sequences and thus increasing the number of cell clones with a correct homologous recombination is the use of a sequence-specific recombination system as described in US 6,110,736, by which unspecifically integrated sequences can be deleted again, which simplifies the selection of events which have integrated successfully via homologous recombination.

[0149] An efficient variant of gene targeting has been reported for *Drosophila melanogaster* (Rong and Golic 2000 Science. 2000 Jun 16;288(5473):2013-8). In this method the construct for targeting is integrated into the host genome flanked by two recognition sites of a site-specific recombinase and includes a site for a rare cutting restriction endonuclease. By induced expression of the site-specific recombinase a DNA circle is excised from the genome. This circle is then linearized after the restriction enzyme (in this case I-SceI) has been expressed resulting in an "activated" DNA molecule with both ends homologous to the target sequence. In the female germ line of *Drosophila,* gene targeting occurred in about one out of 500 cells. Selection of gene targeting events from events of illegitimate recombination can be facilitated by certain combinations of positive and negative selection techniques (WO 99/20780).

[0150] Counter selection is a powerful approach in mammalian and plant systems to enrich for gene targeting events. In plants the bacterial codA gene as a cell autonomous negative selection marker can be used for selection in tissue culture (Schlaman and Hooykaas Plant J 11:1377-1385, 1997; Thykjaer et al., Plant Mol Biol. 1997 Nov;35(4):523-30.). Negative selection in plants allowed a more than a thousand-fold suppression of random integration (Risseeuw et al., Plant J. 1997 Apr;11 (4):717-28. ; Gallego et al., Plant Mol Biol. 1999 Jan;39(1):83-93; Terada et al., Nat Biotechnol. 2002 Oct;20(10):1030-4. Epub 2002 Sep 09.). Exploratory approaches to increase gene targeting in plants comprise expression of proteins like RecA (WO 97/08331) or RecA-homologues derived from other species like e.g., Rad52 (WO 01/68882) or RecA/VirE2 fusion-proteins (WO 01/38504). Use of poly(ADPribose)polymerase inhibitors has demonstrated an increased HR in plants (Puchta H et al. (1995) Plant J 7:203-210). Initiation of sequence-unspecific DNA double-strand breaks was also found to increase efficiency of HR in plants (Puchta H et al. (1995) Plant J 7(2),203-210; Lebel EG et al. (1993) Proc Natl Acad Sci USA 90(2):422-426). However, sequence-unspecific induction of DNA strand breaks is disadvantageous because of the potential mutagenic effect. Sequence-specific induction of DNA strand-breaks may also increase efficiency of HR but is limited to artificial scenarios (Siebert R and Puchta H (2002) Plant Cell 14(5): 1121-1131).

[0151] It is specifically contemplated by the inventors that one could employ techniques for the site-specific integration or excision of transformation constructs prepared in accordance with the instant invention. An advantage of site- specific integration or excision is that it can be used to overcome problems associated with conventional transformation techniques, in which transformation constructs typically randomly integrate into a host genome in multiple copies. This random insertion of introduced DNA into the genome of host cells can be lethal if the foreign DNA inserts into an essential gene. In addition, the expression of a transgene may be influenced by "position effects" caused by the surrounding genomic DNA. Further, because of difficulties associated with plants possessing multiple transgene copies, including gene silencing, recombination and unpredictable inheritance, it is typically desirable to control the copy number of the inserted DNA, often only desiring the insertion of a single copy of the DNA sequence. Site-specific integration or excision of transgenes or parts of transgenes can be achieved in plants by means of homologous recombination (see, for example, U.S. 5,527, 695). The DNA-constructs utilized within the method of this invention may comprise additional nucleic acid sequences. Said sequences may be - for example - localized in different positions with respect to the homology sequences. Preferably, the additional nucleic acid sequences are localized between two homology sequences and may be introduced via homologous recombination into the chromosomal DNA, thereby resembling an insertion mutation of said chromosomal DNA. However, the additional sequences may also be localized outside of the homology sequences (e.g., at the 5'- or 3'-end of the DNA-construct). In cases where the additional sequence resembles a counter selection marker this may allow a distinction of illegitimate insertion events from correct insertion events mediated by homologous recombination. Corresponding negative markers are described below and suitable methods are well known in the art (WO 99/20780).

[0152] In a preferred embodiment of the invention, efficiency of the method of the invention may be further increased by combination with other methods suitable for increasing homologous recombination. Said methods may include for example expression of HR enhancing proteins (like e.g., RecA; WO 97/08331; Reiss B et al. (1996) Proc Natl Acad Sci USA 93(7):3094-3098; Reiss B et al. (2000) Proc Natl Acad Sci USA 97(7):3358-3363) or treatment with PARP inhibitors (Puchta H et al. (1995) Plant J. 7:203-210). Various PARP inhibitors suitable for use within this invention are known to the person skilled in the art and may include for example preferably 3-Aminobenzamid, 8-Hydroxy-2-methylquinazolin-4-on (NU1025), 1,11b-Dihydro-[2H]benzopyrano[4,3,2-de]isoquinolin-3-on (GPI 6150), 5-Aminoisoquinolinon, 3,4-Dihydro-5-[4-(1-piperidinyl) butoxy]-1(2H)-isoquinolinon or compounds described in WO 00/26192, WO 00/29384, WO 00/32579, WO 00/64878, WO 00/68206, WO 00/67734,WO 01/23386 or WO 01/23390. Furthermore, the method may be combined with other methods facilitation homologous recombination and/or selection of the recombinants like e.g., positive/negative selection, excision of illegitimate recombination events or induction of sequence-specific or unspecific

DNA double-strand breaks. In a preferred embodiment, the method for enhancing the expression of a nucleic acid sequence in a plant or a plant cell further via linking the intron with expression enhancing properties to the expression cassette by insertion via homologous recombination is applied to monocotyledonous plants or plant cells, more preferably to plants selected from the group consisting of the genera *Hordeum, Avena, Secale, Triticum, Sorghum, Zea, Saccharum,* and *Oryza,* most preferably a maize plant.

[0153] The nucleic acid sequence in which one of the inventive intron is inserted and functionally linked (via the inventive methods), encodes for a selectable marker protein, a screenable marker protein, a anabolic active protein, a catabolic active protein, a biotic or abiotic stress resistance protein, a male sterility protein or a protein affecting plant agronomic characteristics as described above and/or a sense, antisense, or double-stranded RNA as described above. In a preferred embodiment of the present invention, said nucleic acid sequence encodes a protein. In yet another embodiment of the invention the method is applied to recombinant DNA expression construct that contain a DNA for the purpose of expressing RNA transcripts that function to affect plant phenotype without being translated into a protein. Such non protein expressing sequences comprising antisense RNA molecules, sense RNA molecules, RNA molecules with ribozyme activity, double strand forming RNA molecules (RNAi) as described above.

[0154] Additionally, a further subject matter of the invention relates to the use of the above describes transgenic organism or of cell cultures, parts of transgenic propagation material derived there from, produced with the inventive method, for the production of foodstuffs, animal feeds, seeds, pharmaceuticals or fine chemicals. Preferred is furthermore the use of transgenic organisms for the production of pharmaceuticals or fine chemicals, where a host organism is transformed with one of the above-described expression constructs, and this expression construct contains one or more structural genes which encode the desired fine chemical or catalyze the biosynthesis of the desired fine chemical, the transformed host organism is cultured, and the desired fine chemical is isolated from the culture medium. This process can be used widely for fine chemicals such as enzymes, vitamins, amino acids, sugars, fatty acids, natural and synthetic flavorings, aroma substances and colorants. Especially preferred is the production of tocopherols and tocotrienols, carotenoids, oils, polyunsaturated fatty acids etc. Culturing the transformed host organisms, and isolation from the host organisms or the culture medium, is performed by methods known to the skilled worker. The production of pharmaceuticals such as, for example, antibodies, vaccines, enzymes or pharmaceutically active proteins is described (Hood (1999) Curr Opin Biotechnol. 10(4):382-6;Ma (1999) Curr Top Microbiol. Immunol. 236:275-92; Russel (1999) Current Topics in Microbiology and Immunology 240:119-138; Cramer et al. (1999) Current Topics in Microbiology and Immunology 240: 95-118; Gavilondo (2000) Biotechniques 29(1):128-138 ; Holliger (1999) Cancer & Metastasis Reviews 18(4):411-419).

[0155] Furthermore the present invention relates to recombinant DNA expression construct comprising at least one promoter sequence functioning in plants or plant cells, at least one intron with expression enhancing properties in plants or plant cells characterized by

VIII) an intron length shorter than 1,000 base pairs, and
IX) presence of a 5' splice site comprising the dinucleotide sequence 5'-GT-3' (SEQ ID NO: 78), and
X) presence of a 3' splice site comprising the trinucleotide sequence 5'-CAG-3' (SEQ ID NO: 79), and
XI) presence of a branch point resembling the consensus sequence 5'-CURAY-3' (SEQ ID NO: 75) upstream of the 3'splice site, and
XII) an adenine plus thymine content of at least 40% over 100 nucleotides downstream from the 5' splice site, and
XIII) an adenine plus thymine content of at least 50% over 100 nucleotides upstream from the 3' splice site, and
XIV) an adenine plus thymine content of at least 55%, and a thymine content of at least 30% over the entire intron, and

at least one nucleic acid sequence,
wherein said promoter sequence and at least one of said intron sequences are functionally linked to said nucleic acid sequence and wherein said intron is heterologous to said nucleic acid sequence and/or to said promoter sequence.

Sequences

[0156]

1.	SEQ ID NO: 1	BPSI.1: Sequence of the first intron isolated from the *Oryza sativa* metallothioneine-like gene (accession No. AP002540)
2.	SEQ ID NO: 2	BPSI.2: Sequence of the first intron isolated from the *Oryza sativa* Sucrose UDP Glucosyltransferase-2 gene (accession No. AC084380)
3.	SEQ ID NO: 3	BPSI.3: Sequence of the second intron isolated from the *Oryza sativa* Sucrose UDP Glucosyltransferase-2 gene (accession No. AC084380)

(continued)

4.	SEQ ID NO: 4	BPSI.4: Sequence of the third intron isolated from the *Oryza sativa* Sucrose UDP Glucosyltransferase-2 gene (accession No. AC084380)
5.	SEQ ID NO: 5	BPSI.5: Sequence of the eighth intron isolated from the O. sa *tiva* gene encoding for the Sucrose transporter (accession No. AF 280050).
6.	SEQ ID NO: 6	BPSI.6: Sequence of fourth intron isolated from the *Oryza sativa* gene (accession No. BAA94221) encoding for an unknown protein with homology to the *A. thaliana* chromosome II sequence from clones T22O13, F12K2 encoding for a putative lipase (AC006233).
7.	SEQ ID NO: 7	BPSI.7: Sequence of the fourth intron isolated from the *Oryza sativa* gene (accession No. BAB90130) encoding for a putative cinnamyl-alcohol dehydrogenase.
8.	SEQ ID NO: 8	BPSI.8: Sequence of the second intron isolated from the *Oryza sativa* gene (accession No. AC084766) encoding for a putative ribonucleoprotein.
9.	SEQ ID NO: 9	BPSI.9: Sequence of the fifth intron isolated from the *Oryza sativa* clone GI 12061241.
10.	SEQ ID NO: 10	BPSI.10: Sequence of the third intron isolated from the O. sa *tiva* gene (accession No. AP003300) encoding for a putative protein kinase.
11.	SEQ ID NO: 11	BPSI.11: Sequence of the first intron isolated from the O. sativa gene (accession No. L37528) encoding for a MADS3 box pro tein.
12.	SEQ ID NO: 12	BPSI.12: Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. CB625805) encoding for a putative Adenosylmethionine decarboxylase.
13.	SEQ ID NO: 13	BPSI.13: Sequence of the first intron isolated from the O. *sativa* gene (accession No. CF297669) encoding for an Aspartic proteinase.
14.	SEQ ID NO: 14	BPSI.14: Sequence of the first intron isolated from the O. *sativa* gene (accession No. CB674940) encoding for a Lec14b protein.
15.	SEQ ID NO: 15	BPSI.15: Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. BAD37295.1) encoding for a putative SaIT protein precursor
16.	SEQ ID NO: 16	BPSI.16: Sequence of the first intron isolated from the O. *sativa* gene (accession No. BX928664) encoding for a putative Reticulon.
17.	SEQ ID NO: 17	BPSI.17: Sequence of the first intron isolated from the O. *sativa* gene (accession No. AA752970) encoding for a glycolate oxi dase.
18.	SEQ ID NO: 18	BPSI.18: Sequence of the first intron isolated from the *Oryza sativa* clone (accession No. AK06442 encoding putative non-coding
19.	SEQ ID NO: 19	BPSI.19: Sequence of the first intron isolated from the *Oryza sativa* clone (accession No. AK062197) encoding putative non-coding
20.	SEQ ID NO: 20	BPSI.20 sequence of the first intron isolated from the *O. sativa* gene (accession No. CF279761) encoding for a hypothetical protein.
21.	SEQ ID NO: 21	BPSI.21 Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. CF326058) encoding for a putative membrane transporter.
22.	SEQ ID NO: 22	BPSI.22: Sequence of the firsit intron isolated from the *Oryza sativa* gene (accession No. C26044) encoding for a putative ACT domain repeat protein
23.	SEQ ID NO: 23	Sucrose-UDP glucosyltransferase 2 forward (for) primer
24.	SEQ ID NO: 24	Sucrose-UDP glucosyltransferase 2 reverse (rev) primer
25.	SEQ ID NO: 25	Putative Bowman-Kirk trypsin inhibitor (for) primer
26.	SEQ ID NO: 26	Putative Bowman-Kirk trypsin inhibitor rev primer
27.	SEQ ID NO: 27	Hypothetical protein Acc. No. CF279761 (for) primer
28.	SEQ ID NO: 28	Hypothetical protein Acc. No. CF279761 rev primer
29.	SEQ ID NO: 29	Phenylalanine ammonia-lyase (for) primer
30.	SEQ ID NO: 30	Phenylalanine ammonia-lyase rev primer
31.	SEQ ID NO: 31	Metallothioneine-like protein 1 (for) primer
32.	SEQ ID NO: 32	Metallothioneine-like protein 1 rev primer
33.	SEQ ID NO: 33	Catalase (for) primer
34.	SEQ ID NO: 34	Catalase rev primer
35.	SEQ ID NO: 35	Putative stress-related protein (for) primer

(continued)

36.	SEQ ID NO: 36	Putative stress-related protein rev primer
37.	SEQ ID NO: 37	Putative translation initiation factor SUI1 (for) primer
38.	SEQ ID NO: 38	Putative translation initiation factor SUI1 rev primer
39.	SEQ ID NO: 39	Polyubiquitin (for) primer
40.	SEQ ID NO: 40	Polyubiquitin rev primer
41.	SEQ ID NO: 41	Glutathione S-transferase II (for) primer
42.	SEQ ID NO: 42	Glutathione S-transferase II rev primer
43.	SEQ ID NO: 43	Metallothioneine-like protein 2 (for) primer
44.	SEQ ID NO: 44	Metallothioneine-like protein 2 rev primer
45.	SEQ ID NO: 45	Translational initiation factor eIF1 (for) primer
46.	SEQ ID NO: 46	Translational initiation factor eIF1 rev primer
47.	SEQ ID NO: 47	OSJNBa0024F24.10 (unknown protein) (for) primer
48.	SEQ ID NO: 48	OSJNBa0024F24.10 (unknown protein) rev primer
49.	SEQ ID NO: 49	Protein similar to Histone 3.2-614 (for) primer
50.	SEQ ID NO: 50	Protein similar to Histone 3.2-614 rev primer
51.	SEQ ID NO: 51	OSJNBa0042L16.3 (for) primer
52.	SEQ ID NO: 52	OSJNBa0042L16.3 rev primer
53.	SEQ ID NO: 53	BPSI.1-5' primer
54.	SEQ ID NO: 54	BPSI.1-3' primer
55.	SEQ ID NO: 55	BPSI.2-5' primer
56.	SEQ ID NO: 56	BPSI.2-3' primer
57.	SEQ ID NO: 57	BPSI.3-5' primer
58.	SEQ ID NO: 58	BPSI.3-3' primer
59.	SEQ ID NO: 59	BPSI.4-5' primer
60.	SEQ ID NO: 60	BPSI.4-3' primer
61.	SEQ ID NO: 61	BPSI.5-5' primer
62.	SEQ ID NO: 62	BPSI.5-3' primer
63.	SEQ ID NO: 63	BPSI.6-5' primer
64.	SEQ ID NO: 64	BPSI.6-3' primer
65.	SEQ ID NO: 65	BPSI.7-5' primer
66.	SEQ ID NO: 66	BPSI.7-3' primer
67.	SEQ ID NO: 67	BPSI.8-5' primer
68.	SEQ ID NO: 68	BPSI.8-3' primer
69.	SEQ ID NO: 69	BPSI.9-5' primer
70.	SEQ ID NO: 70	BPSI.9-3' primer
71.	SEQ ID NO: 71	BPSI.10-5' primer
72.	SEQ ID NO: 72	BPSI.10-3' primer
73.	SEQ ID NO: 73	BPSI.11-5' primer
74.	SEQ ID NO: 74	BPSI.11-3' primer
75.	SEQ ID NO: 75	5'-CURAY-3' plant branchpoint consensus sequences 1
76.	SEQ ID NO: 76	5'-YURAY-3' plant branchpoint consensus sequences 2
77.	SEQ ID NO: 77	5'-(AG)(AG)/GT(AGT)(AGT)(GTC)-3' preferred 5' splice-site
78.	SEQ ID NO: 78	5'splice site dinucleotide 5'-GT-3'
79.	SEQ ID NO: 79	3'splice site trinucleotide 5'-CAG-3'
80.	SEQ ID NO: 80	5'splice site plant consensus sequence 5'-AG::GTAAGT-3'
81.	SEQ ID NO: 81	3' splice site plant consensus sequence 5'-CAG::GT-3'
82.	SEQ ID NO: 82	Sequence of the first intron isolated from the *Oryza sativa* metallothioneine-like gene (accession No. AP002540) including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.1 (SEQ ID NO:1)

(continued)

83.	SEQ ID NO: 83	Sequence of the first intron isolated from the O. *sativa* Sucrose UDP Glucosyltransferase-2 gene (accession No. AC084380) including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.2 (SEQ ID NO:2)
84.	SEQ ID NO: 84	Sequence of the second intron isolated from the O. *sativa* Sucrose UDP Glucosyltransferase-2 gene (accession No. AC084380) including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.3 (SEQ ID NO:3)
85.	SEQ ID NO: 85	Sequence of the third intron isolated from the O. *sativa* Sucrose UDP Glucosyltransferase-2 gene (accession No. AC084380) including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.4 (SEQ ID NO:4)
86.	SEQ ID NO: 86	Sequence of the eighth intron isolated from the *Oryza sativa* gene encoding for the Sucrose transporter (GenBank accession No. AF 280050) including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.5 (SEQ ID NO:5)
87.	SEQ ID NO: 87	Sequence of the eighth intron isolated from the *Oryza sativa* gene encoding for the Sucrose transporter (accession No. AF 280050) including modified 5' and 3' splice sites and sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI. 5 (SEQ ID NO:5)
88.	SEQ ID NO: 88	Sequence of the fourth intron isolated from the *Oryza sativa* gene encoding for an unknown protein with homology to the *A.thaliana* chromosome II sequence from clones T22O13, F12K2 encoding for a putative lipase (AC006233) including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.6 (SEQ ID NO:6)
89.	SEQ ID NO: 89	Sequence of the fourth intron isolated from the *Oryza sativa* gene encoding for an unknown protein with homology to the *A.thaliana* chromosome II sequence from clones T22O13, F12K2 encoding for a putative lipase (AC006233) including modified 5' and 3' splice sites and sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI. 6 (SEQ ID NO:6)
90.	SEQ ID NO: 90	Sequence of the fourth intron isolated from the *Oryza sativa* gene (accession No. BAB90130) encoding for a putative cinnamyl-alcohol dehydrogenase including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.7 (SEQ ID NO:7)
91.	SEQ ID NO: 91	Sequence of the fourth intron isolated from the *Oryza sativa* gene (accession No. BAB90130) encoding for a putative cinnamyl-alcohol dehydrogenase including modified 5' and 3' splice sites and sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.7 (SEQ ID NO:7)
92.	SEQ ID NO: 92	Sequence of the second intron isolated from the *Oryza sativa* gene (accession No. AC084766) encoding for a putative ribonucleoprotein including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.8 (SEQ ID NO:8)
93.	SEQ ID NO: 93	Sequence of the second intron isolated from the *Oryza sativa* gene (accession No. AC084766) encoding for a putative ribonucleoprotein including modified 5' and 3' splice sites and sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.8 (SEQ ID NO:8)
94.	SEQ ID NO: 94	Sequence of the third intron isolated from the *Oryza sativa* gene (accession No. AP003300) encoding for a putative protein including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.10 (SEQ ID NO:10)
95.	SEQ ID NO: 95	Sequence of the third intron isolated from the *Oryza sativa* gene (accession No. AP003300) encoding for a putative protein including modified 5' and 3' splice sites and sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.10 (SEQ ID NO:10)
96.	SEQ ID NO: 96	Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. L37528) encoding for a MADS3 box protein including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.11 (SEQ ID NO:11)
97.	SEQ ID NO: 97	Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. L37528) encoding for a MADS3 box protein including modified 5' and 3' splice sites and sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.11 (SEQ ID N0:11)

(continued)

98.	SEQ ID NO: 98	Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. CB625805) encoding for a putative Adenosylmethionine decarboxylase including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.12 (SEQ ID NO:12)
99.	SEQ ID NO: 99	Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. CF297669) encoding for a Aspartic proteinase including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.13 (SEQ ID NO:13)
100.	SEQ ID NO: 100	Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. CB674940) encoding for a Lec14b protein including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.14 (SEQ ID NO:14)
101.	SEQ ID NO: 101	Sequence of the first intron isolated from the O. *sativa* gene (accession No. CA128696) encoding for a putative mannose-binding rice lectin including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.15 (SEQ ID NO: 15)
102.	SEQ ID NO: 102	Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. BX928664) encoding for a putative Reticulon including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.16 (SEQ ID NO:16)
103.	SEQ ID NO: 103	Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. AA752970) encoding for a glycolate oxidase including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.17 (SEQ ID NO:17)
104.	SEQ ID NO: 104	Sequence of the first intron isolated from the *Oryza sativa* clone GI 34763855 including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.18 (SEQ ID NO:18)
105.	SEQ ID NO: 105	Sequence of the first intron isolated from the *Oryza sativa* clone GI 32533738 including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.19 (SEQ ID NO:19)
106.	SEQ ID NO: 106	Sequence of the first intron isolated from the *Oryza sativa* gene (accession No. CF279761) encoding for a hypothetical protein including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.20 (SEQ ID NO:20).
107.	SEQ ID NO: 107	Sequence of the first intron isolated from the O. *sativa* gene (accession No. CF326058) encoding for a putative membrane transporter including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.21 (SEQ ID NO:21).
108.	SEQ ID NO: 108	Sequence of the first intron isolated from the O. *sativa* gene (accession No. C26044) encoding for a putative ACT domain repeat protein including sequences 5' and 3' adjacent to the 5' and 3' splice sites of the intron sequence BPSI.22 (SEQ ID NO:22).
109.	SEQ ID NO: 109	Binary vector pBPSMM291
110.	SEQ ID NO: 110	Binary vector pBPSMM305
111.	SEQ ID NO: 111	Binary vector pBPSMM350
112.	SEO ID NO: 112	Binary vector pBPSLM139
113.	SEQ ID NO: 113	Artificial sequence: cassette from vector pBPSMM355 (OsCP12::BPSI.1) comprising Os CP12 promoter (bp 1 - 854) and BPSI.1 intron (bp 888 - 1470).
114.	SEQ ID NO: 114	Artificial sequence: cassette from from vector pBPSMM355 (ZmHRGP::BPSI.1) comprising Maize [HRGP] hydroxyproline-rich glycoprotein (extensin) 5'/UTR promoter (bp 1 - 1184) and oryza sativa BPSI.1 intron (bp 1217- 1799)
115.	SEQ ID NO: 115	Artificial sequence: cassette from vector pBPSMM358 (OsCCoAMT1::BPS1.1) comprising p-caffeoyl-CoA 3-O-methyltransferase [CoA-O-Methyl] promoter (bp 1 - 1034) and BPS1.1 intron (1119 - 1701)
116.	SEQ ID NO: 116	Artificial sequence: cassette from vector EXS1025 (ZmGlobulinl::BPSI.1) comprising Maize Globulin-1 [ZmGlb1] promoter (W64A) (bp 1 - 1440)and BPSI.1 intron (1443 - 1999)
117.	SEQ ID NO: 117	Artificial sequence: cassette from vector pBPSMM369 (OsV-ATPase::BPSI.1)comprising putative Rice H+-transporting ATP synthase 5'/UTR promoter (1 - 1589) and BPSI.1 intron (1616 - 2198)
118.	SEQ ID NO: 118	Artificial sequence: cassette from vector pBPSMM366 (OsC8,7SI::BPSI.1) comprising Putative Rice C-8,7 Sterol isomerase promoter (1 - 796) and BPSI.1 intron (827 - 1409)

(continued)

119.	SEQ ID NO: 119	Artificial sequence: cassette from vector pBPSMM357 (ZmLDH::BPSI.1) comprising maize gene Lactate Dehydrogenase 5'/UTR promoter (bp 1 - 1062) and BPSI.1 intron (bp 1095 - 1677).
120.	SEQ ID NO: 120	Artificial sequence: cassette from vector pBPSLM229 (ZmLDH::BPSI.5) comprising maize gene Lactate Dehydrogenase 5'/UTR promoter (bp 1 - 1062) and BPSI.5 intron (bp 1068 - 1318)
121.	SEQ ID NO: 121	Artificial sequence: cassette from vector pBPSMM371 (OsLea::BPSI.1)comprising rice Lea (Late Embryogenesis Abundant) promoter (bp 1 - 1386) and BPSI.1 intron (bp 1387 - 2001)

EXAMPLES

Chemicals

[0157] Unless indicated otherwise, chemicals and reagents in the Examples were obtained from Sigma Chemical Company (St. Louis, MO), restriction endonucleases were from New England Biolabs (Beverly, MA) or Roche (Indianapolis, IN), oligonucleotides were synthesized by MWG Biotech Inc. (High Point, NC), and other modifying enzymes or kits regarding biochemicals and molecular biological assays were from Clontech (Palo Alto, CA), Pharmacia Biotech (Piscataway, NJ), Promega Corporation (Madison, WI), or Stratagene (La Jolla, CA). Materials for cell culture media were obtained from Gibco/BRL (Gaithersburg, MD) or DIFCO (Detroit, MI). The cloning steps carried out for the purposes of the present invention, such as, for example, restriction cleavages, agarose gel electrophoresis, purification of DNA fragments, transfer of nucleic acids to nitrocellulose and nylon membranes, linking DNA fragments, transformation of *E. coli* cells, growing bacteria, multiplying phages and sequence analysis of recombinant DNA, are carried out as described by Sambrook (1989). The sequencing of recombinant DNA molecules is carried out using ABI laser fluorescence DNA sequencer following the method of Sanger (Sanger 1977).

Example 1: Identification and characterization of IME-introns in highly expressing genes

1.1 Identification of strongly and constitutively expressed *Oryza sativa* gene candidates.

[0158] Using the above described "sequencing by hybridization method" *in silico* clone distribution analysis of rice cDNA libraries have been performed.
The rice cDNA clone distribution profiles were derived from about 7.6 million rice cDNA clones, which were generated over 23 rice cDNA libraries of different tissues at different developmental stages and biotic/abiotic treatments. Method for the production of cDNA libraries are well known in the art (*e.g.* Gubler U, and Hoffman BJ. (1983) A simple and very efficient method for generating cDNA libraries. Gene 25(2-3):263-269.; Jung-Hwa Oh et al. (2003) An improved method for constructing a full-length enriched cDNA library using small amounts of total RNA as a starting material. EXPERIMENTAL and MOLECULAR MEDICINE 35(6):586-590; Lanfranchi et al. (1996) Identification of 4370 expressed sequence tags from a 3'-end-specific cDNA library of human skeletal muscle by DNA sequencing and filter hybridization. Genome Res. 6(1):35-42). Furthermore, a comprehensive description of cDNA library construction is provided in 1) Cowell and Austin. cDNA Library Protocols. In Methods in Molecular Biology, Volume 69, October 1996, Humana Press, Scientific and medical publishers, ISBN: 0-89603-383-X; and 2) Ying, Shao-Yao. Generation of cDNA Libraries, Methods and Protocols. In Methods in Molecular Biology, Volume 221, February 2003, Humana Press, Scientific and medical publishers, ISBN: 1-58829-066-2.
[0159] All of the clones were clustered into a total of 300,408 rice clusters using the above described (see "sequencing by hybridization method", or "HySeq-technology") high-throughput technology of 288 plant-specific 7 mer-oligonucleotide fingerprinting. For each generated cluster, clones have further been clustered into different variants using more stringent cutoff value of the hybridization pattern similarity, leading to 335,073 rice clone variants. Therefore, within each variant for given cluster, clones are more homogeneous. The distribution of rice cDNA clones over the 23 normalized cDNA libraries for given variants provides the rice variant expression profiles. The normalized cDNA library was produced by first adjusting the orignal library clone size to the average clone size of all of the 23 libraries, then adjusting the number of clones per variant in that library based on the adjusted total number of clones in that library. Rice clones are selected from the rice clusters for sequencing to generate rice EST data. In using the clones distribution profiles of 335,073 rice variants, 145 variants were selected based on the number of clones exceeding top 1 % of the clone distribution across the entire library for over each of 23 libraries, and genes were identified using the homologs to the EST sequences

derived from the variants. These candidate genes showed strong, constitutive, and ubiquitous expression. The rice EST sequences homolog to these candidate genes were mapped to the rice genomic DNA sequences. Top 15 candidates out of 145 were selected based on availability of genomic sequences, annotation, and high level of expression (Table 2).

Table 2. Gene candidates for potential IME-introns

Candidate gene	Annotation
1	sucrose-UDP glucosyltransferase 2
2	putative Bowman-Kirk trypsin inhibitor
3	Hypothetical Protein
4	phenylalanine ammonia-lyase
5	metallothioneine-like protein1
6	Catalase
7	putative stress-related protein
8	putative translation initiation factor SUI1
9	Polyubiquitin
10	glutathione S-transferase II
11	metallothioneine-like protein2
12	translational initiation factor eIF1
13	OSJNBa0024F24.10 (Unknown Protein)
14	Similar to Histone 3.2-614
15	OSJNBa0042L16.3

1.2 Validation of highly expressing gene candidates using real time RT-PCR

[0160] Expression levels of the endogenous genes representing these 15 candidates were measured at the mRNA levels using LightCycler. Total RNA was extracted from rice plants at various developmental stages and tissues with and without drought stress (6, 12, 24, and 48 hr by withholding water) using Qiagen RNeasy Plant Mini Kit (Cat. No 74904). Quality and quantity of the RNA were determined using Molecular Probes RiboGreen Kit (Cat. No. R-11490) on the Spectra MAX Gemini. One μg of RNA was used for RT-PCR (Roche RT-PCR AMV kit, Cat. No. 1483188) in the reaction solution I (1 μg RNA, 2 μL 10x Buffer, 4 μL 25 mM $MgCl_2$, 2 μL 1 mM dNTPs, 2 μL 3.2 μg Random Primers, 1 μL 50 units RNase Inhibitor, 0.8 μL 20 units AMV-RT polymerase, fill to 20 μL with sterile water) under the optimized PCR program (25°C 10 min, 42°C 1 hr, 99°C 5 min, 4°C stop reaction).
The RT-PCR samples were used for the LightCycler reaction (11.6 μL sterile water, 2.4 μL 25mM $MgCl_2$, 2 μL SYBR Green Polymerase mix, 2 μL 10mM Specific Primer Mix, 2 μL RT-PCR reaction product) under the optimized program (95°C 5 min, 95°C 30 sec, 61 °C 40 sec, 72°C 40 sec and repeat steps 2-4 for 30 cycles, 72°C 10 min, and 4°C stop reaction) provided by Roche (LightCycler FastStart DNA Master SYBR Green I, Cat. No.3003230).

Table 3. Primer sequences of the gene candidates

Gene	Primers	SEQ ID NO.
Sucrose-UDP glucosyltransferase 2	Fwd: 5'-tttgtgcagcccgctttctacgag Rev: 5'-acggccaacgggacggtgcta	23 24
Putative Bowman-Birk trypsin inhibitor	Fwd: 5'-gtcctcgccggcatcgtcac Rev: 5'-cagaacggcgggttgatcc	25 26
Hypothetical protein Acc. No. CF279761	Fwd: 5'-agctcgctcgcggtctt Rev: 5'-acagggcccaagtcgtgtgc	27 28
Phenylalanine ammonia-lyase	Fwd: 5'-aggtctcgccatcgtcaatg Rev: 5'-cgagacgggcgttgt	29 30

(continued)

Gene	Primers	SEQ ID NO.
Methallothioneine-like protein 1	Fwd: 5'-ggctgcggaggatgcaagatg Rev: 5'-ggggttgcaggtgcagttgtcg	31 32
Catalase	Fwd: 5'-ggcgtcaacacctacacctt Rev: 5'-tgcactgcagcatcttgtcgtc	33 34
Putative stress-related protein	Fwd: 5'-ggtggatgccacggtgcaagag Rev: 5'-ggggaggtactgtgctc	35 36
Putative translation initiation factor SUI1	Fwd: 5'-tgcggaagccaatgctga Rev: 5'-ccagccctgaactaggaacgtc	37 38
Polyubiquitin	Fwd: 5'-tcaggggaggcatgcaaa Rev: 5'-tgcataccaccacggagacgaa	39 40
Glutathione S-transferase II	Fwd: 5'-cgatttctccaaaggcgagcac Rev: 5'-tgcgggtatgcgtccaaca	41 42
Metallothioneine-like protein 2	Fwd: 5'-acagccaccaccaagaccttcg Rev: 5'-ctgcagctggtgccacacttgc	43 44
Translational initiation factor eIF1	Fwd: 5'-tcccaactgccttcgatccctt Rev: 5'-tggacagtggtcaggctcttacgg	45 46
OSJNBa0024F24.10 (unknown protein)	Fwd: 5'-gagttctaccagttcagcgacc Rev: 5'-aacccgaaggcgttgac	47 48
Similar to Histone 3.2-614	Fwd: 5'-agaccgcccgcaagtc Rev: 5'-cttgggcatgatggtgacgc	49 50
OSJNBa0042L16.3	Fwd: 5'-ccaagagggagtgctgtatgccaa Rev: 5'-acgaggaccaccacggtacccat	51 52

[0161] Standardizing the concentration of RNA (1 μg) in each of the RT-PCR reactions was sufficient to directly compare the samples if the same primers were used for each Lightcycler reaction. The output results were a number that corresponds to the cycle of PCR at which the sample reaches the inflection point in the log curve generated. The lower the cycle numbers the higher the concentration of target RNA present in the sample. Each sample was repeated in triplicate and an average was generated to produce the sample "crosspoint" value. The lower the crosspoint, the stronger the target gene was expressed in that sample. (*Roche Molecular Biochemicals* LightCycler System: Reference Guide May 1999 version) Based on the LightCycler results, 11 candidates were selected (Table 4).

Table 4. LightCycler results representing expression of the rice gene candidates at the mRNA levels.

Gene candidates [strong & constitutive expression]	Drought stressed rice root (R) and shoot (S) (hr withholding water)								Well-watered conditions			
	R6	R12	R24	R48	S6	S12	S24	S48	seedling	Panicle during flowering stage	shoots	flowers
Unknown	21.1	21.6	N/A	20.3	20.5	21.7	N/A	21.0	23.3	22.7	21.4	23.7
Catalase	21.2	22.7	26.7	26.0	21.9	21.7	N/A	27.8	22.8	31	20.6	23.5
GSTII	20.6	20.3	23.3	23.7	21.8	23.2	N/A	20.6	24.4	22.6	22.1	24.8
Hypothetical Protein	31	31	31	31	31	31	31	31	31	31	27.4	27.0
Metallothioneine 1	20.1	21.5	16.5	16.3	18.3	19.8	N/A	19.2	21.0	22.5	20.6	20.6
Metallothioneine2	20.2	20.8	23.8	24.8	18.5	18.7	N/A	18.7	19.9	17.8	21.2	19.2
PolyUbuiquitin	19.5	19.1	19.4	20.4	19.1	20.4	N/A	19.8	22.8	20.7	20.0	22.6
Stress Related Protein	24.1	23.9	23.7	24.0	23.4	23.4	N/A	23.3	24.6	24.0	23.6	24.9
Sucrose-UDP glucoryltransferase 2	21.3	21.9	26.6	26.7	20.7	20.9	27.2	22.6	20.9	19.1	20.7	26.0
SUI1	21.3	21.1	23.1	23.6	21.9	22.8	N/A	21.7	24.4	23.8	22.9	30.2
TIF	23.6	23.6	N/A	22.9	22.1	23.3	N/A	23.1	24.6	23.8	22.8	23.7
Trypsin Inhibitor	24.0	23.8	24.5	25.0	22.8	23.3	23.5	23.2	26.2	23.8	23.2	23.05

The numbers represent PCR cycle that reaches the start of the exponential curve of the PCR product. Lower the number indicates that higher the expression of the endogenous gene is.

1.3 Identification of IME-introns

[0162] Candidate introns were isolated using the public available genomic DNA sequences (e.g. http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html), leading to a total of 20 introns, mostly first, second, and/or third introns from the targeted genes. These intron sequences were screened by the following IME criteria:

- 5' splice site GT, 3' splice site CAG
- At least 40% AT rich over 100 nucleotides downstream from the 5' splice site GT
- At least 50% AT rich over 100 nucleotides upstream from the 3' splice site CAG
- At least 55% AT rich and 35% T rich over the entire intron
- CURAY branch point
- Intron size less than 1kb

[0163] Selected intron candidates can retain up to 50 bp exon sequences upstream and downstream of the 5' and 3' splice sites, respectively.

[0164] After screening the intron sequences against the IME criteria described above, four out of the 20 candidates were chosen and named as follows.

Table 5. The intron candidates

Intron name	Annotation
BPSI.1 (SEQ ID No. 1)	Metallothioneine1 first intron
BPSI.2 (SEQ ID No. 2)	Sucrose-UDP glucosyltransferase2 first intron
BPSI.3 (SEQ ID No. 3)	Sucrose-UDP glucosyltransferase2 second intron
BPSI.4 (SEQ ID No. 4)	Sucrose-UDP glucosyltransferase2 third intron

1.4 Isolation of the intron candidates

[0165] Genomic DNA from rice was extracted using the Qiagen DNAeasy Plant Mini Kit (Qiagen). Genomic DNA regions containing introns of interest were isolated using conventional PCR. Approximately 0.1 μg of digested genomic DNA was used for the regular PCR reaction (see below). The primers were designed based on the rice genomic sequences. One μL of the diluted digested genomic DNA was used as the DNA template in the primary PCR reaction. The reaction comprised six sets of primers (Table 6) in a mixture containing Buffer 3 following the protocol outlined by an Expand Long PCR kit (Cat #1681-842, Roche-Boehringer Mannheim). The isolated DNA was employed as template DNA in a PCR amplification reaction using the following primers:

Table 6. Primer sequences

Primer name	Sequence
BPSI.1-5' (SEQ ID No. 53)	5'-cccgggcaccctgcggagggtaagatccgatcacc
BPSI.1-3' (SEQ ID No. 54)	5'-cggaccggtacatcttgcatctgcatgtac
BPSI.2-5' (SEQ ID No. 55)	5'-cccgggcacccttcaccaggttcgtgctgatttag
BPSI.2-3' (SEQ ID No. 56)	5'-cggaccgaaccagcctgcgcaaataacag
BPSI.3-5' (SEQ ID No. 57)	5'-cccgggcacctcctgaggagtgcacaggtttg
BPSI.3-3' (SEQ ID No. 58)	5'-cggaccgggagataacaatcccctcctgcatg
BPSI.4-5' (SEQ ID No. 59)	5'-cccgggcacccagcttgtggaagaagggtatg
BPSI.4-3' (SEQ ID No. 60)	5'-cggaccggttgttggtgctgaaatatacatc

[0166] Amplification was carried out in the PCR reaction (5 μL 10X Advantage PCR Mix [Eppendorf], 5 μL genomic

DNA [corresponds to approximately 80 ng], 2.5 mM of each dATP, dCTP, dGTP and dTTP [Invitrogen: dNTP mix], 1 μL of 20 μM 5'-intron specific primer 20pM, 1 μL of 20 μM 3' intron specific primer, 1 μL TripleMaster DNA Polymerase mix [Eppendorf], in a final volume of 50 μL) under the optimized PCR program (1 cycle with 15 sec at 94°C and 1 min at 80°C 35cycles with 15 sec at 94°C, 1 min at 58°C and 1 min at 72°C) provided by Thermocycler (T3 Thermocycler Biometra).

[0167] The PCR product was applied to an 1% (w/v) agarose gel and separated at 80V. The PCR products were excised from the gel and purified with the aid of the Qiagen Gel Extraction Kit (Qiagen, Hilden, Germany). The PCR product can be cloned directly into vector pCR4-TOPO (Invitrogen) following the manufacturer's instructions, *i.e.* the PCR product obtained was inserted into a vector having T overhangs with its A overhangs and a topoisomerase.

1.5 Vector Construction

[0168] The base vector to which the intron candidates were clone in was pBPSMM267. This vector comprises the maize ubiquitin promoter with no intronic sequence, followed by multiple cloning sites (MCS) to be used for addition of introns of interest, then the GUSint ORF (including the potato invertase [PIV]2 intron to prevent bacterial expression), followed by nopaline synthase (NOS) terminator. The intron-containing expression vectors were generated by ligation of XmaI-RsrII digested intron PCR products into XmaI-*Rsr*II linearized pBPSMM267, thereby resulting in the following vectors (Table 7).

Table 7. GUS chimeric constructs containing introns in the 5' UTR

pUC-based expression vector	Binary vector	Composition of the expression cassette (promoter::intron::reporter gene::terminator)
pBPSMM291	pBPSMM350	Zm.ubiquitin promoter::BPSI.1::GUS::NOS3'
pBPSMM293	pBPSMM353	Zm.ubiquitin promoter::BPSI.2::GUS::NOS3'
pBPSMM294	pBPSMM312	Zm.ubiquitin promoter::BPSI.3::GUS::NOS3'
pBPSMM295	pBPSMM310	Zm.ubiquitin promoter::BPSI.4::GUS::NOS3'

1.6 Plant analysis for identifying IME-introns

[0169] These experiments were performed by bombardment of plant tissues or culture cells (Example 4.1), or by Agrobacterium-mediated transformation (Example 4.3). The target tissues for these experiments can be plant tissues (e.g. leaf or root), cultured cells (e.g. maize BMS), or plant tissues (e.g. immature embryos) for *Agrobacterium* protocols.

1.6.1 Transient assays

[0170] To identify IME-introns, four introns (BPSI.1, 2, 3, and 4) were tested using Microprojectile bombardment. The maize ubiquitin promoter (Zm.ubiquitin) without any intronic sequence was used as basal expression (negative control). Introns of interest were cloned into the 5'UTR region of Zm.ubiquitin promoter. Maize ubiquitin intron was used as a positive control to measure the relative levels of expression enhanced by introns of interest based on GUS expression. Strong enhancement with BPSI.1 and BPSI.2 introns was detected (Table 8). BPSI.3 intron showed medium enhancement levels of GUS expression. No expression was detected with BPSI.4 intron.

Table 8. Transient GUS expression testing for intron-mediated enhancement

Intron candidates	GUS expression*	
Zm.ubiquitin promoter alone (negative control)	++	50%**
Zm.ubiquitin promoter + Zm.ubiquitin intron1 (positive control)	++++	100%
Zm.ubiquitin promoter+ BPSI.1 (pBPSMM291)	++++	100%
Zm.ubiquitin promoter + BPSI.2 (pBPSMM293)	++++	100%
Zm.ubiquitin promoter + BPSI.3 (pBPSMM294)	+++	80%

(continued)

Intron candidates	GUS expression*	
Zm.ubiquitin promoter + BPSI.4 (pBPSMM295)	-	0%
*GUS histochemical assays: a range of GUS activities (- no expression to ++++ high expression), **Relative GUS expression compared to the expression controlled by maize ubiquitin promoter fused with Zm.ubiquitin intron.		

1.6.2 Analysis of IME-intron candidates in stably transformed maize

[0171] The binary vectors pBPSMM350, pBPSMM353, pBPSMM312, and pBPSMM310 (Table 7), were transformed into maize using Agrobacterium-mediated transformation (Example 4.3). The levels and patterns of GUS expression controlled by BPSI.1, BPSI.2, BPSI.3, or BPSI.4 intron were compared with those controlled by Zm.ubiquitin intron. BPSI.1, BPSI.2 and BPSI.3 introns enhanced expression in roots, leaves, and kernels throughout the various development stages at a similar level to that observed in transient assays (Table 9). Expression of Zm.ubiquitin promoter without intron was undetectable in roots and leaves and was limited in kernels to the endosperm. Expression of Zm.ubiqutin promoter with BPSI.4 intron exhibited the same expression patterns as those controlled by Zm.ubiquitin promoter without intron. This result indicates that a transient assay can be used as a model system and is therefore one of the important screening systems to identify introns that function in intron-mediated enhancement (IME) in stable transformed plants. However, the results obtained with the transient assays should be validated by the production of stable transformed transgenic plants.

Table 9. GUS expression in transgenic maize plants

Developmental stage	Organs	Zmubiquitin promoter:: Zmu biquitin intron	Zmubiquitin promoter:: no intron	Zmubiquitin promoter:: BPSI.1 (pPSMM350)
Five leaf	Roots	++++	-	++++
	Leaves	++++	-	+++
Flowering	Leaves	++++	-	+++
Late reproductive	Kernels	++++	++**	+++
Developmental stage	Organs	Zmubiquitin promoter:: BPSI.2 (pBPSMM353)	Zmubiquitin promoter:: BPSI.3 (pBPSMM312)	Zmubiquitin promoter:: BPSI.4 (pBPSMM310)
Five leaf	Roots	+++	+++	-
	Leaves	+++	++	-
Flowering	Leaves	+++	+++	-
Late reproductive	Kernels	+++	+++	++**
*GUS histochemical assays: a range of GUS activities (- no expression to ++++ high expression), ** only in endosperm, ND: not determined				

EXAMPLE 2. IME-introns located in the annotated DNA sequences

2.1 *In silico* screening system

[0172] The *in silico* intron-screening system for identifying introns that have the functional IME comprises three major components: (1) Generate intron sequence database and screen for intron candidates using the functional IME criteria (indicated in Example 1.3); (2) Define the expression profiles of these candidate genes from which introns were selected; (3) Further examine the selected gene structures by conducting a mapping of EST sequences onto the genomic region where the candidate genes resided.

[0173] More than 30,000 annotated rice and maize genomic sequences were downloaded from NCBI. Intron, 5'- and

3'-UTR, promoter and terminator sequences were isolated (*in silico*) from those annotated genes and their corresponding sequence databases were generated (Table 10, 11). From the generated intron sequence database, more than 111,800 introns (*i.e.,* 106049 rice introns, 4587 maize introns) were screened for potential intron regulatory enhancement elements based on the functional IME criteria (see 1.3). A total of 108 potential intron candidates have been identified, and the protein sequences of the intron candidate genes were retrieved from NCBI. The rice (we do not disclose maize sequences) homolog EST sequences were identified from the cDNA libraries described in example 1 using the BLASTx algorithm (this program compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against protein sequences) at an E-value of $1.0e^{-20}$ against those protein sequences. Using the rice variant expression profiling data (see example 1), the introns whose genes were homolog to the rice genes with desirable expression profiling, such as constitutive and tissue specific expression pattern, were selected as final *in silico* identified intron candidates for lab experimental test.

[0174] The rice UniGenes, which was derived from the EST sequence assembly, were updated using the combined public rice EST data and the EST data obtained using the databases described in example 1, and the UniGene expression profiling data was generated using the rice variant expression profiling data over the 23 different libraries described in example 1. The newly updated rice UniGene expression profiling data were used to help select the final 108-intron candidates. Perl scripts have been written to isolate intron, 5'- and 3'-UTR, terminator, and promoter sequences from the entire NCBI rice and maize annotated gnomic DNA sequences for creating corresponding sequence databases, to screen for functional IME, and to compare the expression profiling data (see example 5). The introns were retrieved from the CDS (coding sequences) features of the annotated genes. A total of 106,049 rice introns and 4,587 maize introns have been retrieved (Table 10) from more that 30,000 annotated genes as the data summarized in Table 11 and 12.

Table 10. Rice/maize sequence database summary

	Rice	Maize
Intron	106049	4587
5' UTR	129	236
3' UTR	142	694
Terminator	7	5
Promoter	69	239

Table 11. Rice and maize gene summary*

Average	Rice	Maize
gene length	2471	3223
intron length	399	279
extron length	309	388
intron/gene	3.9	2.61
extron/gene	4	2.45
GC/intron	39%	40.8%
GC/extron	54.8%	55.3%
* Intron or extron without gene names were excluded from the calculation.		

Table 12. Total number of genes in the database

Species	Gene Name	Gene Identifier
Rice	30059	30249
Maize	1281	3549

[0175] Furthermore, The full length coding sequences of all 108 candidate genes, in which introns were isolated, were downloaded from NCBI and blasted against the Hyseq rice and maize UniGenes to identify Hyseq rice and maize homolog sequences, using BLASTN and $1.0e^{-20}$ cutoff E-value. Top hits of rice UniGenes were selected, and the gene expression profiling data was examined. The EST sequences, identified as homolog to the coding sequences of selected intron candidate genes, were retrieved and mapped along with the intron candidate gene sequences to the rice genomic regions. Based on the UniGene expression profiling data and the candidate gene structures, annotated and confirmed by the EST sequence alignments, nine introns were finally selected from a total of 108 intron candidates and are subject to the real time RT-PCR expression test. Among the nine introns, four showed a constitutive expression pattern, three preferably expressed in the early seed-developed stage, one preferably expressed in root, and one was induced in the drought condition (Table 13).

Table 13. Intron candidates selected based on the second *in silico* screening system

Intron	Rice GI number	Sequence homology
BPSI.5 (SEQ ID No. 5)	9624451	Sucrose transporter
BPSI.6 (SEQ ID No. 6)	7523493	Similar to Arabidopsis thaliana chromosome II sequence from clones T22O13, F12K2; putative lipase (AC006233)
BPSI.7 (SEQ ID No. 7)	20161203	putative cinnamyl-alcohol dehydrogenase
BPSI.8 (SEQ ID No. 8)	18921322	Putative ribonucleoprotein
BPSI.9 (SEQ ID No. 9)	12061241	putative mitochondrial carrier protein
BPSI.10 (SEQ ID No. 10)	20160990	Putative protein kinase
BPSI.11 (SEQ ID No. 11)	886404	5'UTR intron (1^{st}) MADS3 box protein

2.2 Isolation of the intron candidates

[0176] Genomic DNA from rice was extracted using the Qiagen DNAeasy Plant Mini Kit (Qiagen). Genonic DNA regions containing introns of interest were isolated using conventional PCR. Approximately 0.1 μg of digested genomic DNA was used for the regular PCR reaction (see below). The primers were designed based on the rice genomic sequences. Five μL of the diluted digested genomic DNA was used as the DNA template in the PCR reaction. PCR was performed using the TripleMaster PCR System (Eppendorf, Hamburg, Germany) as described by the manufacturer.

Table 14. Primers used for amplification of widely expressed intron candidates

Primers	Sequence
BPSI.5-5' (SEQ ID No. 61)	5'-cggggtaccgagctctctggtggctgaggtaagttctgttattacc
BPSI.5-3' (SEQ ID No. 62)	5'-cggggatccggacaggaaaacctgaaaacaggg
BPSI.6-5' (SEQ ID No. 63)	5'-cggggtaccgagctcgacgatttaggtaagtcattattgtctc
BPSI.6-3' (SEQ ID No. 64)	5'-cggggatcctcactgaaacctgcagtgtagg
BPSI.7-5' (SEQ ID No. 65)	5'-cggggtaccgagctcgatcctaaggtaagcactctagctg
BPSI.7-3' (SEQ ID No. 66)	5'-cggggatccgtaactcaacctgttttttta
BPSI.8-5' (SEQ ID No. 67)	5'-cggggtaccgagctccaatggctaggtaagtatatgcttcc
BPSI.8-3' (SEQ ID No. 68)	5'-cggggatcccccatcaagtacctgttttaag
BPSI.9-5' (SEQ ID No. 69)	5'-cggggtaccgagctcgaatacctaggtaagtccatctc
BPSI.9-3' (SEQ ID No. 70)	5'-cggggatcccacacaagcgacctggaaaaataagc
BPSI.10-5' (SEQ ID No. 71)	5'-cggggtaccgagctcccatcttttaggtaagtatctttgcg
BPSI.10-3' (SEQ ID No. 72)	5'-cggggatccggtaaagaacctgtttaatac
BPSI.11-5' (SEQ ID No. 73)	5'-cggggtaccgagctctgaacaggaaggtaagttctggctttcttgc
BPSI.11-3' (SEQ ID No. 74)	5'-cggggatcctcagatcgacctggacacaaacgc

[0177] Amplification was carried out in the PCR reaction (5 μL 10X Advantage PCR Mix [Eppendorf], 5 μL genomic DNA [corresponds to approximately 80 ng], 2.5 mM of each dATP, dCTP, dGTP and dTTP [Invitrogen: dNTP mix], 1 μL of 20 μM 5'-intron specific primer 20pM, 1 μL of 20 μM 3' intron specific primer, 1 μL TripleMaster DNA Polymerase mix [Eppendorf], in a final volume of 50 μL) under the optimized PCR program (1 cycle with 15 sec at 94°C and 1 min at 80°C 35cycles with 15 sec at 94°C, 1 min at 58°C and 1 min at 72°C) provided by Thermocycler (T3 Thermocycler Biometra).

[0178] A QIAspin column was used to purify the PCR products as directed by the manufacturer (Qiagen, Valencia, CA), and the amplified introns were used directly for cloning into expression vectors, as described below.

2.3 Vector Construction

[0179] The base expression vector for these experiments was pBPSMM305, which comprises the maize lactate dehydrogenase (LDH) promoter without intron driving expression of the GUSint gene followed by the NOS terminator. The LDH promoter has been demonstrated to direct undetectable levels of GUS expression by colorimetric staining in the absence of an intron capable of providing IME.

[0180] Intron PCR products were digested with *Sac*I & *Bam*HI and cloned into pBPSMM305 linearized with *Sac*I & *Bam*HI, generating the following LDH:intron:GUS expression vectors.

Table 15. GUS chimeric constructs containing introns in the 5' UTR

pUC-based expression vector	Composition of the expression cassette (promoter::intron::reporter gene::terminator)
pBPSJB041 (pBPSLI017)	ZmLDH promoter::BPSI.5::GUS::NOS3'
pBPSJB042 (pBPSLI018)	ZmLDH promoter::BPSI.6::GUS::NOS3'
pBPSJB043 (pBPSLI019)	ZmLDH promoter::BPSI.7::GUS::NOS3'
pBPSJB044 (pBPSLI020)	ZmLDH promoter::BPSI.8::GUS::NOS3'
pBPSJB045 (pBPSLI021)	ZmLDH promoter::BPSI.9::GUS::NOS3'
pBPSJB046 (pBPSLI022)	ZmLDH promoter::BPSI.10::GUS::NOS3'
pBPSJB050 (pBPSLI023)	ZmLDH promoter::BPSI.11::GUS::NOS3'

[0181] Binary vector pBPSLI017 comprises the expression cassette containing the BPSI.5 intron and was generated by ligating in the *Pme*I-*Pac*I fragment from pBPSJB041 into pBPSLM139 linearized with *Pme*I and *Pac*I.

[0182] Binary vector pBPSLI018 comprises the expression cassette containing the BPSI.6 intron and was generated by ligating in the *Pme*I-*Pac*I fragment from pBPSJB042 into pBPSLM139 linearized with *Pme*I and *Pac*I.

[0183] Binary vector pBPSLI019 comprises the expression cassette containing the BPSI.7 intron and was generated by ligating in the *Pme*I-*Pac*I fragment from pBPSJB043 into pBPSLM139 linearized with *Pme*I and *Pac*I.

[0184] Binary vector pBPSLI020 comprises the expression cassette containing the BPSI.8 intron and was generated by ligating in the *Pme*I-*Pac*I fragment from pBPSJB044 into pBPSLM139 linearized with *Pme*I and *Pac*I.

[0185] Binary vector pBPSLI021 comprises the expression cassette containing the BPSI.9 intron and was generated by ligating in the *Pme*I-*Pac*I fragment from pBPSJB045 into pBPSLM139 linearized with *Pme*I and *Pac*I.

[0186] Binary vector pBPSLI022 comprises the expression cassette containing the BPSI.10 intron and was generated by ligating in the *Pme*I-*Pac*I fragment from pBPSJB046 into pBPSLM139 linearized with *Pme*I and *Pac*I.

[0187] Binary vector pBPSLI023 comprises the expression cassette containing the BPSI.11 intron and was generated by ligating in the *Pme*I-*Pac*I fragment from pBPSJB050 into pBPSLM139 linearized with *Pme*I and *Pac*I.

2.4 Transient assays for identifying the intron functioning IME

[0188] These experiments were performed by bombardment of plant tissues or culture cells (Example 4.1), or by Agrobacterium-mediated transformation (Example 4.3). The target tissues for these experiments can be plant tissues (e.g. leaf or root), cultured cells (*e.g.* maize BMS), or plant tissues (e.g. immature embryos) for *Agrobacterium* protocols. Characterization of these introns for their ability to direct IME in conjunction with the LDH promoter was undertaken via transient expression by bombardment of expression vectors into maize leaf tissue and liquid-cultured BMS cells, respectively.

[0189] The maize lactate dehydrogenase promoter (ZmLDH) without any intronic sequence was used as basal expression (negative control). Introns of interest were cloned into the 5'UTR region of ZmLDH promoter. Maize ubiquitin

intron was used as a positive control to measure the relative levels of expression enhanced by introns of interest based on GUS expression.

Due to the very low background (no detectable GUS expression) of the ZmLDH promoter in the absence of intron, the presence of any GUS staining indicates that a particular intron is capable of providing IME. Of the introns tested, BPSI.10 and BPSI.11 introns consistently yielded the highest GUS expression, at a level comparable to the LDH::Zm.ubiquitin intron construct. In addition to these introns, BPSI.5, BPSI.6, and BPSI.7 introns consistently resulted in an intermediate level of GUS expression in between LDH alone and LDH::Zm.ubiquitin intron. Comparable results were obtained in maize leaves and BMS cells, indicating that the tested introns confer IME in green and non-green tissues (Table 16).

Table 16. Transient GUS expression testing for intron-mediated enhancement

Intron candidates	GUS expression*	
	leaves	BMS
No intron (Zm.LDH promoter alone)	-	-
Zm.LDH + Zm.ubiquitin intron (positive control)	++++	++++
Zm.LDH promoter + BPSI.5	++	++
Zm.LDH promoter + BPSI.6	+++	+++
Zm.LDH promoter + BPSI.7	+++	+++
Zm.LDH promoter + BPSI.8	-	+
Zm.LDH promoter + BPSI.9	-	-
Zm.LDH promoter + BPSI.10	++++	+++
Zm.LDH promoter + BPSI.11	++++	ND
*GUS histochemical assays: a range of GUS activities (- no expression to ++++ high expression), ND: not determined.		

EXAMPLE 3. Identification of IME-introns located in the 5' untranslated region

3.1 *In silico* screening system

[0190] The *in silico* intron screening system for identifying introns that have the functional IME located in the '5 UTR comprises three major components: (1) Genome mapping of the entire rice CDS, released from Institute of Genome Research on October 2, 2003 and the EST sequence collections; (2) identification and selection of the introns located in the 5'UTR using both the functional IME criteria and the rice cDNA clone distribution profiles; (3) validation of the selected 5'UTR introns by examining the sequence alignments among the genomic DNA, CDS and ESTs, the gene model, sequence reading frame and intron splicing sites

[0191] A total of 56,056 annotated rice CDS were mapped onto the Japonica rice genome in which both rice CDS and genomic DNA sequences were obtained from The Institute of Genome Research. Additional 422,882 rice EST sequences of public and in-house sources were also mapped onto the rice genome. A splicing alignment software, GeneSeqer (version September 2, 2003 from Iowa State University Research foundation), was used to conduct the entire genome mapping. Since both EST and CDS were mapped onto their corresponding genomic regions, the sequence alignment coordinators [coordinators are the start and/or end positions of the the genomic sequences where CDS/EST sequences aligned to] derived from the CDS mapping and the EST mapping on the same genomic region provide opportunity to identify the alignment extension of the EST sequences along the genomic DNA beyond the start codon of the CDS. Such sequence alignment extension from the EST sequences beyond CDS indicates the identification of the 5' UTRs, which have not been contained in the CDS, but in the EST sequences. The system selects these EST sequences, which extend the sequence alignment beyong the CDS along the gnome for up to 5k base long for 5'URT intron screening. For any predicted exons, the last exon in the prediceted 5'UTR region must aligned at the same position of the 1st exon of the CDS. The gnome mapping results have identified 461 genes that have their 5' UTR containing at least one intron.

[0192] Further stringent screen criteria that required at least 3 EST sequences confirming the same predicted 5'UTR introns were used to select the gene candidates, leading to identify 87 gene candidates. Those identified EST sequences, which were considered as the same transcript as the rice CDS, were used to retrieve the rice cDNA clone distribution data or the microarray expression data in which either the clones of those identified EST sequences have been spotted on the rice microarray chip or homolog to those identified EST sequences were identified on the chip. For given the rice

cDNA clone distribution profile, a gene, which has a cluster/variant size of more than 100 clones distributed over 23 cDNA libraries, was considered highly expressed. For given the microarray expression, a gene, which has hybridization signal intensity exceeding the top 25% percentile within the same sample, was also considered highly expressed. In addition to the gene expression criteria used for gene candidate selection, the IME criteria (indicated in Example 1.3) were applied.

[0193] Furthermore, a validation of the selected candidate genes was conducted by examining the coincidence of the sequence alignments between EST, CDS sequences and genomic DNA sequence. Clearly the EST sequences needed to support the gene model predicted from the CDS. Any conflict of the sequence alignments between EST and CDS would result in the deselecting the candidate genes. Using those criteria, a final list of 11 introns was selected (Table 17).

Table 17. Intron candidates selected based on the third *in silico* screening system

Intron	Rice GI number	Sequence homology
BPSI.12 (SEQ ID No. 12)	29620794	Putative adenosylmethionine decarboxylase
BPSI.13 (SEQ ID No. 13)	33666702	Aspartic proteinase
BPSI.14 (SEQ ID No. 14)	29678665	Lec14b protein
BPSI.15 (SEQ ID No. 15)	35009827	Putative mannose-binding rice lectin
BPSI.16 (SEQ ID No. 16)	41883853	Putative reticulon
BPSI.17 (SEQ ID No. 17)	2799981	Glycolate oxidase
BPSI.18 (SEQ ID No. 18)	34763855	Similar to AT4g33690/T16L1_180
BPSI.19 (SEQ ID No. 19)	32533738	N/A
BPSI.20 (SEQ ID No. 20)	33657147	Hypothetical protein
BPSI.21 (SEQ ID No. 21)	33800379	Putative membrane transporter
BPSI.22 (SEQ ID No. 22)	2309889	Putative ACT domain repeat protein

3.2 Isolation of introns

[0194] Genomic DNA containing introns of interest is isolated using conventional PCR amplification with sequence specific primers (see 1.4) followed by cloning into a PCR cloning vector in the art.

3.3 Vector construction

[0195] Introns are PCR amplified from rice genomic DNA using primers that engineer a *Sac*I site on the 5' end of the intron and a *Bam*HI site on the 3' end of the sequence. The PCR products are digested with *Sac*I and *Bam*HI and ligated into pBPSMM305 linearized with *Sac*I and *Bam*HI to generate pUC-based expression vectors comprising the Zm.LDH promoter::Intron candidate::GUSint::NOS terminator.

[0196] Binary vectors for stable maize transformation are constructed by digesting the pUC expression vectors with *Pme*I and *Pac*I and ligating into pBPSLM139 digested with *Pme*I and *Pac*I.

3.4 Transient assays for identifying IME-introns

[0197] These experiments are performed by bombardment of plant tissues or culture cells (Example 4.1), or by Agrobacterium-mediated transformation (Example 4.3). The target tissues for these experiments can be plant tissues (*e.g.* leaf or root), cultured cells (*e.g.* maize BMS), or plant tissues (*e.g.* immature embryos) for *Agrobacterium* protocols.

EXAMPLE 4. Assays for identifying IME-introns

[0198] These experiments are performed by bombardment of plant tissues or culture cells (Example 4.1), by PEG-mediated (or similar methodology) introduction of DNA to plant protoplasts (Example 4.2), or by Agrobacterium-mediated transformation (Example 4.3). The target tissue for these experiments can be plant tissues (*e.g.* leaf tissue), cultured plant cells (*e.g.* maize Black Mexican Sweetcorn (BMS), or plant embryos for *Agrobacterium* protocols.

4.1 Transient assay using microprojectile bombardment

[0199] The plasmid constructs are isolated using Qiagen plasmid kit (cat# 12143). DNA is precipitated onto 0.6 μM gold particles (Bio-Rad cat# 165-2262) according to the protocol described by Sanford *et al.* (1993) and accelerated onto target tissues (e.g. two week old maize leaves, BMS cultured cells, etc.) using a PDS-1000/He system device (Bio-Rad). All DNA precipitation and bombardment steps are performed under sterile conditions at room temperature.

[0200] Black Mexican Sweet corn (BMS) suspension cultured cells are propagated in BMS cell culture liquid medium [Murashige and Skoog (MS) salts (4.3 g/L), 3% (w/v) sucrose, myo-inositol (100 mg/L), 3 mg/L 2,4-dichlorophenoxyacetic acid (2,4-D), casein hydrolysate (1 g/L), thiamine (10mg/L) and L-proline (1.15 g/L), pH 5.8]. Every week 10 mL of a culture of stationary cells are transferred to 40 mL of fresh medium and cultured on a rotary shaker operated at 110 rpm at 27°C in a 250 mL flask.

[0201] 60 mg of gold particles in a siliconized Eppendorf tube are resuspended in 100% ethanol followed by centrifugation in a Mini centrifuge C1200 (National Labnet Co. Wood-bridge, NJ) for 30 seconds. The pellet is rinsed once in 100% ethanol and twice in sterile water with centrifugation after each wash. The pellet is finally resuspended in 1 mL sterile 50% glycerol. The gold suspension is then divided into 50 μL aliquots and stored at 4°C. The following reagents are added to one aliquot: 5 μL of 1 μg/pL total DNA, 50 μL 2.5M $CaCl_2$, 20μL 0.1M spermidine, free base. The DNA solution is vortexed for 1 minute and placed at -80°C for 3 min followed by centrifugation for 10 seconds in a Mini centrifuge C1200. The supernatant is removed. The pellet is carefully resuspended in 1 mL 100% ethanol by flicking the tube followed by centrifugation for 10 seconds. The supernatant is removed and the pellet is carefully resuspended in 50 μL of 100% ethanol and placed at -80°C until used (30 min to 4 hr prior to bombardment). If gold aggregates are visible in the solution the tubes are sonicated for one second in a waterbath sonicator just prior to use.

[0202] For bombardment, two-week-old maize leaves are cut into pieces approximately 1 cm in length and placed adaxial side up on osmotic induction medium M-N6-702 [N6 salts (3.96 g/L), 3% (w/v) sucrose, 1.5 mg/L 2,4-dichlorophenoxyacetic acid (2,4-D), casein hydrolysate (100 mg/L), and L-proline (2.9 g/L), MS vitamin stock solution (1 mL/L), 0.2 M mannitol, 0.2 M sorbitol, pH 5.8]. The pieces are incubated for 1-2 hours.

[0203] In the case of BMS cultured cells, one-week-old suspension cells are pelleted at 1000 g in a Beckman/Coulter Avanti J25 centrifuge and the supernatant is discarded. Cells are placed onto round ash-free No 42 Whatman filters as a 1/16 inch thick layer using a spatula. The filter papers holding the plant materials are placed on osmotic induction media at 27°C in darkness for 1-2 hours prior to bombardment. Just before bombardment the filters are removed from the medium and placed onto on a stack of sterile filter paper to allow the calli surface to partially dry.

[0204] Each plate is shot with 6 μL of gold-DNA solution twice, at 1,800 psi for the leaf materials and at 1,100 psi for the BMS cultured cells. To keep the position of plant materials, a sterilized wire mesh screen is laid on top of the sample. Following bombardment, the filters holding the samples are transferred onto M-N6-702 medium lacking mannitol and sorbitol and incubated for 2 days in darkness at 27°C prior to transient assays. Transient expression levels of the reporter genes are determined by GUS staining, quantification of luminescence or RT-PCR using the protocols in the art. GUS staining is done by incubating the plant materials in GUS solution [100 mM NaHPO4, 10 mM EDTA, 0.05% Triton X100, 0.025% X-Gluc solution (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid dissolved in DMSO), 10% methanol, pH 7.0] at 37°C for 16-24 hours. Plant tissues are vacuum-infiltrated 2 times for 15 minutes to aid even staining.

[0205] Transient expression levels of the reporter genes are determined by staining, enzyme assays or RT-PCR using the protocols in the art.

4.2 Transient assay using protoplasts

[0206] Isolation of protoplasts is conducted by following the protocol developed by Sheen (1990). Maize seedlings are kept in the dark at 25°C for 10 days and illuminated for 20 hours before protoplast preparation. The middle part of the leaves are cut to 0.5 mm strips (about 6 cm in length) and incubated in an enzyme solution containing 1% (w/v) cellulose RS, 0.1% (w/v) macerozyme R10 (both from Yakult Honsha, Nishinomiya, Japan), 0.6 M mannitol, 10 mM Mes (pH 5.7), 1 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM β-mercaptoethanol, and 0.1 % BSA (w/v) for 3 hr at 23°C followed by gentle shaking at 80 rpm for 10 min to release protoplasts. Protoplasts are collected by centrifugation at 100 x g for 2 min, washed once in cold 0.6 M mannitol solution, centrifuged, and resuspended in cold 0.6 M mannitol (2 x 10^6/mL).

[0207] A total of 50 μg plasmid DNA in a total volume of 100 μL sterile water is added into 0.5 mL of a suspension of maize protoplasts (1 x 10^6 cells/mL) and mix gently. 0.5 mL PEG solution (40 % PEG 4,000, 100 mM $CaNO_3$, 0.5 mannitol) is added and pre-warmed at 70°C with gentle shaking followed by addition of 4.5 mL MM solution (0.6 M mannitol, 15 mM $MgCl_2$, and 0.1 % MES). This mixture is incubated for 15 minutes at room temperature. The protoplasts are washed twice by pelleting at 600 rpm for 5 min and resuspending in 1.0 mL of MMB solution [0.6 M mannitol, 4 mM Mes (pH 5.7), and brome mosaic virus (BMV) salts (optional)] and incubated in the dark at 25°C for 48 hr. After the final wash step, collect the protoplasts in 3 mL MMB medium, and incubate in the dark at 25°C for 48 hr. Transient expression levels of the reporter gene are determined quantification of expression of reporter genes or RT-PCR using the protocols

in the art in order to determine potentially intron candidates that function in intron-mediated enhancement.

4.3 Agrobacterium-mediated transformation in dicotyledonous and monocotyledonous plants

4.3.1 Transformation and regeneration of transgenic *Arabidopsis thaliana* (Columbia) plants

[0208] To generate transgenic *Arabidopsis* plants, *Agrobacterium tumefaciens* (strain C58C1 pGV2260) is transformed with the various vector constructs described above. The *Agrobacterial* strains are subsequently used to generate transgenic plants. To this end, a single transformed *Agrobacterium* colony is incubated overnight at 28°C in a 4 mL culture (medium: YEB medium with 50 μg/mL kanamycin and 25 μg/mL rifampicin). This culture is subsequently used to inoculate a 400 mL culture in the same medium, and this is incubated overnight (28°C, 220 rpm) and spun down (GSA rotor, 8,000 rpm, 20 min). The pellet is resuspended in infiltration medium (1/2 MS medium; 0.5 g/L MES, pH 5.8; 50 g/L sucrose). The suspension is introduced into a plant box (Duchefa), and 100 ml of SILWET L-77 (heptamethyltrisiloxan modified with polyalkylene oxide; Osi Specialties Inc., Cat. P030196) is added to a final concentration of 0.02%. In a desiccator, the plant box with 8 to 12 plants is exposed to a vacuum for 10 to 15 minutes, followed by spontaneous aeration. This is repeated twice or 3 times. Thereupon, all plants are planted into flowerpots with moist soil and grown under long-day conditions (daytime temperature 22 to 24°C, nighttime temperature 19°C; relative atmospheric humidity 65%). The seeds are harvested after 6 weeks.

[0209] As an alternative, transgenic *Arabidopsis* plants can be obtained by root transformation. White root shoots of plants with a maximum age of 8 weeks are used. To this end, plants that are kept under sterile conditions in 1 MS medium (1% sucrose; 100mg/L inositol; 1.0 mg/L thiamine; 0.5 mg/L pyridoxine; 0.5 mg/L nicotinic acid; 0.5 g MES, pH 5.7; 0.8 % agar) are used. Roots are grown on callus-inducing medium for 3 days (1x Gamborg's B5 medium; 2% glucose; 0.5 g/L mercaptoethanol; 0.8% agar; 0.5 mg/L 2,4-D (2,4-dichlorophenoxyacetic acid); 0.05 mg/L kinetin). Root sections 0.5 cm in length are transferred into 10 to 20 mL of liquid callus-inducing medium (composition as described above, but without agar supplementation), inoculated with 1 mL of the above-described overnight *Agrobacterium* culture (grown at 28°C, 200 rpm in LB) and shaken for 2 minutes. After excess medium has been allowed to run off, the root explants are transferred to callus-inducing medium with agar, subsequently to callus-inducing liquid medium without agar (with 500 mg/L betabactyl, SmithKline Beecham Pharma GmbH, Munich), incubated with shaking and finally transferred to shoot-inducing medium (5 mg/L 2-isopentenyladenine phosphate; 0.15 mg/L indole-3-acetic acid; 50 mg/L kanamycin; 500 mg/L betabactyl). After 5 weeks, and after 1 or 2 medium changes, the small green shoots are transferred to germination medium (1 MS medium; 1% sucrose; 100 mg/L inositol; 1.0 mg/L thiamine; 0.5 mg/L pyridoxine; 0.5 mg/L nicotinic acid; 0.5 g MES, pH 5.7; 0.8% agar) and regenerated into plants.

4.3.2 Transformation and regeneration of crop plants

[0210] The *Agrobacterium*-mediated plant transformation using standard transformation and regeneration techniques may also be carried out for the purposes of transforming crop plants (Gelvin& Schilperoort (1995) Plant Molecular Biology Manual, 2nd Edition, Dordrecht: Kluwer Academic Publ. ISBN 0-7923-2731-4; Glick & Thompson (1993) Methods in Plant Molecular Biology and Biotechnology, Boca Raton: CRC Press, ISBN 0-8493-5164-2). For example, oilseed rape can be transformed by cotyledon or hypocotyl transformation (Moloney (1989) Plant Cell Reports 8: 238-242). The use of antibiotics for the selection of agrobacteria and plants depends on the binary vector and the *Agrobacterium* strain used for the transformation. The selection of oilseed rape is generally carried out using kanamycin as selectable plant marker. The Agrobacterium-mediated gene transfer in linseed (Linum usitatissimum) can be carried out using for example a technique described by Mlynarova (1994) Plant Cell Report 13:282-285. The transformation of soybean can be carried out using, for example, a technique described in EP-A1 0 424 047 or in EP-A1 0 397 687, US 5,376,543, US 5,169,770. The transformation of maize or other monocotyledonous plants can be carried out using, for example, a technique described in US 5,591,616. The transformation of plants using particle bombardment, polyethylene glycol-mediated DNA uptake or via the silicon carbonate fiber technique is described, for example, by Freeling & Walbot (1993) "The maize handbook" ISBN 3-540-97826-7, Springer Verlag New York).

EXAMPLE 5: Computer algorithm for retrieving sequence information from NCBI genebank file.

[0211] The target feature keys are intron, terminator, promoter, UTR. The following script (written in computer language Pearl) is giving an example for a computer algorithm of the invention suitable to identify suitable intron sequences based of database information (see also Fig. 5a-f):

```
#!/usr/local/bin/perl -w
# intron.pl
```

```
open(IN,$ARGV[0]) or die "can't find output";

while (defined(my $file=<IN> )) {

#start of a single annotation

  if ($file=~/LOCUS.*?\s+(\d+)\sbp(.*)/) {
           my $length=$1;
           my $mol=1;
          $mol=0 if $2 =~ /circular/;
        my @cdslist=();
        my @start=();
           my $order=0; # order=1: complementary coding.
        my @title=();
        my @title0=();
        my @intron=();
        my $id="";
        my @terminator=();
        my @promoter=();
        my @utr5=();
        my @utr3=();
        my @origin=();
        my $tab="";
        my $organism="";
           while (defined(my $line=<IN> )) {
          $line=$tab.$line;

          if ($line =~ /^VERSION.*?\s+(GI:\d+)/) {
                  $id=$1;

          }elsif ($line =~ /^\s{2}ORGANISM\s+(.*)/) {

             if($1=~/Oryza sativa/i) {
                $organism="rice";
                 }elsif($1=~/Zea mays/i) {
                $organism="maize";
             }elsif($1=~/Glycine max/i) {
                $organism="soybean";
             }else {
                $1=~/(\w+)/;
            $organism=$1;
             }
          }elsif($line =- /^\s{5} (CDS\s*)/) { #extract cds
               my $test=$';
               my $gene="N/A";
           my $start=1;
           my $product="N/A";
           my $gi=$id;
           my @cds=();
           my @temp=();
           if ($test =- /complement/) {
                    $order=1 ;
               }else {
                    $order = 0;
               }
               while ( my $in=<IN>) {
             if ($in =- /\s\/(.*)/) {
                  $test=$test;
                  if ($1=~/gene="(.*)"/) {
                       $gene=$1;
                  }elsif($1=~/note="(.*)"/) {
                       $product=$1;
```

```
                    }else {
                    last;
                    }
                } else {
                    $test=$test.$in;
                }
            } #close while loop;

                $test =~s/\w+\d+\.\d:\d+\.\.\d+//g;
                $test =- s/\D/ /g;
                    $test =~ s/\s+/ /g;
                    $test =~ s/^\s+//;
                my @sort;
            if ($mol==0) {
                @sort=split(/ /,$test);
            } else {
                    @sort=sort {$a <=> $b} split(/ /,$test);
            }
# tag complement cds
            if ($order==1) {
                    @cds = ("complement",@sort);
                } elsif ($order==0) {
                    @cds = @sort;
                } #close if loop;
#retreave notation if intron exist;
                if (scalar(@cds) >= 4) {

                   while (my $in=<IN>) {
                      $start=1;
                         if ($in =~ /codon_start=(\d+)/) {
                      $start = $1;
                    }elsif ($in =~ /\/gene="(.*)"/) {
                      $gene=$1;
                      }elsif ($in =~ /\/product=(.*)/) {
                      $product=$1;
                      $Product=~ tr/'"'//d;
                      }elsif ($in=~ /db_xref="(GI:.*?)"/) {
                      $gi = $1;

                      last;
                          } elsif ($in=~ /\/(pseudo)/) {
                      $product="pseudo";
                    last;
                    } #close if loop
                  } #close while loop;
                  push @start, $start;
                  push @cdslist, \@cds;
# retreave 5'utr if start codon > 1;
                  my @tem=();
                  for (my $i=1;$i<=($#cds-1)/2;$i++) {
                    my $title1=">$organism|$gi|Intron_$i ";
                    my $title2=" $gene|$start|".($cds[2*$i-
1+$order]+1)."..".($cds[2*$i+$order]-1)."|$product\n";
                    my @title=($title1,$title2);
                    push @tem, \@title;
                  } #close for loop
                  push @title, \@tem;
                  my $title0=">$organism|$gi|5UTR_0
$gene|$start|".($cds[$order]-1)."..".($cds[$order]+$start-
2)."|$product\n";
                  push @title0, $title0;

                  } #close if @cds>4 loop
```

```
                    } elsif ($line =- /^\s{5}terminator/) {

                      ($tab,my $note,my @term)=&getTerminator($line);

                  push @terminator, $note;
                  push @terminator, \@term;

                 } elsif ($line =- /^\s{5}promoter/) {

                      ($tab,my $note,my @prom)=&getTerminator($line);

                     push @promoter, $note;
                  push @promoter, \@prom;

                } elsif ($line =- /^\s{5}5\DUTR/) {

                      ($tab,my $note,my @temp)=&getTerminator($line);

                  push @utr5,$note;
                  push @utr5,\@temp;

              } elsif ($line =- /^\s{5}3\DUTR/) {

                     ($tab,my $note,my @temp)=&getTerminator($line);

                  push @utr3,$note;
                  push @utr3,\@temp;

        #get sequence @origin
                }
               if ($line =- /^(ORIGIN)/) {
                     $line="";
                  while (my $code=<IN>) {

                    if ($code =- /\/\//) {
                          last;
                       }else{
                       $line=$line.$code;
                    } #close if loop
                      } #close while loop
                  # $line =- s/\/\// /g;
                # print $line,"\n";
                           $line =~ tr/0-9//d;
                           $line =~ tr/ //d;
                           $line =~ tr/\n//d;
                     @origin = split(//,$line);
                for (my $i=0; $i<=$#cdslist;$i++) {

                 if ($start[$i]>2) {

                     my @first=();
                     my $first;
                     if (${$cdslist[$i]}[0] eq "complement") {
                       my @utr=@origin[$cdslist[$i][1]-1 ..
($cdslist[$i][1]+$start[$i]-2)];
                   print @utr,"\n";
                       $first=&complement(@utr);
                     } else {
                        @first=@origin[$cdslist[$i][0]-1 ..
($cdslist[$i][0]+$start[$i]-2)];
                     $first=join(",@first);
                     } #close if loop for complement
```

```
                print $title0[$il,$first,"\n\n";
                } #close if loop for $start>2;

          if (${$cdslist[$i]}[0] eq "complement") {
              shift @{$cdslist[$i]};
             for (my $j=1; $j<=($#{$cdslist[$i]}-1)/2;$j++) {
                my @int=@origin[$cdslist[$i][2*$j-1] ..
$cdslist[$i][2*$j]-2];
           my $intl=&complement(@int);
           print $title[$i][$j-1][0],scalar(@int),$title[$i][$j-
1][1], $int1,"\n\n" if $#int<5000;
            } #close 2nd for loop for complement

           } else {
              for (my $j=1; $j<=($#{$cdslist[$i]}-1)/2;$j++) {
           my @int=@origin[$cdslist[$i][2*$j-1] .. $cdslist[$i][2*$j]-

2];
           if ($mol==0 && $cdslist[$i][2*$j-1] > $cdslist[$i][2*$j]) {
               @int=(@origin[$cdslist[$i][2*$j-1] .. $#origin],
@origin[0 .. $cdslist[$i][2*$j]-2]);
               }
           my $int1=join(",@int);
               print $title[$i][$j-1][0],scalar(@int),$title[$i][$j-
1][1], $int1,"\n\n" if $#int < 5000;
         }#close 2nd for loop
               } #close else loop
            } #close 1st for loop

         my $title1=">$organism|$id|terminator";
         &getSequence(\@terminator,\@origin,$titlel);
            $title1=">$organism|$id|promoter";
         &getSequence(\@promoter,\@origin,$title1);

            $title1=">$organism|$id|5utr";
         &getSequence(\@utr5,\@origin,$title1);
            $title1=">$organism|$id|3utr";
         &getSequence(\@utr3,\@origin,$title1);
       last;

      } else {
        $tab="";
          } #close if $line loop
       } #close while $line loop
       next;

    } #close if $file loop

 } #close while $file loop
close IN;

#retreave complement sequnce
sub complement {
 my @code=@_;
 my @complemnt=();
 for (my $i=0;$i<=$#code;$i++) {
    if ($code[$#code-$i] eq "t") {
        $complement[$i]= "a";
    } elsif ($code[$#code-$i] eq "a") {
        $complement[$i]= "t";
    } elsif ($code[$#code-$i] eq "c") {
        $complement[$i]= "g";
    } elsif ($code[$#code-$i] eq "g") {
```

```
     $complement[$i]= "c";
    } else {
        $complement[$i]=$code[$#code-$i];
    }#close if loop
 } #close for loop
 my $comp=join(",@complement);
 @complement=();
 return $comp;
} #close sub

#get sequence reference for feature keys
sub getTerminator {
my $line=$_[0];
my $order=0;
if ($line=~/complement/) {
     $order=1;
} else {
} #close if loop
$line =~ s/\d'UTR//;
$line =~ s/\D/ /g;
$line =~ s/\s+/ /g;
$line =~ S/A\S//;
my @term=split(' ',$line);
    @term=("c",@term) if $order==1;
my $in;
read(IN,$in,6);
my $note =" \n";

if ($in!~/\w/) {
    $note=<IN>;
    $note=~s/\s+\///;
    $note=~s/note=//;
    $note=~ tr/""//d;
} #close if loop
return ($in,$note,@term);
} #close sub

#retreave sequence information for feature keys
sub getSequence {
my @array=@{$_[0]};
my @code=@{$_[1]};
my $id=$_[21;

for (my $i=0; $i<($#array+1)/2;$i++) {
      my $note=$array[2*$i];
      my @term=@{$array[2*$i+1]};
    if ($term[0] eq "c") {
      shift @term;
      for (my $j=0; $j<=($#term-1)/2;$j++) {
        my @comp=@code[($term[2*$j]-1) .. ($term[2*$j+1]-1)];
        my $int1=&complement(@comp);
        my $title=$id."_".($i+1)." ".scalar(@comp)."
$term[2*$j]..$term[2*$j+1]$note";
        print $title, $int1,"\n\n";
      } #close 2nd for loop
    } else {
      for (my $j=0; $j<($#term+1)/2;$j++) {
    my @int=@code[($term[2*$j]-1) .. ($term[2*$j+1]-1)];
        my $int1=join('',@int);
        my $title=$id."_".($i+1)." ".scalar(@int)."
$term[2*$j]..$term[2*$j+1]|$note";
    print $title, $int1,"\n\n";
      } #close 2nd for loop
```

```
    } #close if loop
  } #close 1st for loop
 } #close sub
```

EXAMPLE 6 Expression of tissue-specific promoters in combination with IME-introns

[0212] BPSI.1 and BPSI.5 have been fused with various monocot promoters and demonstrated that most of these promoters without IME-intron did not show GUS expression, but IME-introns have enhanced expression.

6.1 Os.CP12 promoter::BPSI.1 intron::GUS::NOS terminator (pBPSMM355)

[0213] pBPSMM355 shows strong leaf-specific expression. This expression was detected in all tested developmental stages. No expression was detected in any other tissue tested.

6.2 Zm.HRGP promoter::BPSI.1 intron::GUS::NOS terminator (pBPSMM370)

[0214] pBPSMM370 is strongly expressed in roots. Significant expression was also detected in silk and in the outermost layers of the kernel that include the aleuron layer and seed coat. This expression was strongest around the base of the kernel. Staining in silk was strongest in the region close to the attachment point with the kernel and was detected at very early developmental stages.

6.3 Os.CCoAMT1 promoter::BPSI.1 intron::GUS::NOS terminator (pBPSMM358)

[0215] Os.Caffeoyl-CoA-O-methyltransferase (CCoAMT1) promoter in combination with BPSI.1 (pBPSMM358) showed embryo-specific expression in T1 and T2 kernels. The expression level was low but very specific. No expression was detected in any other tissue tested.

6.4 Zm.Globulin1 promoter::BPSI.1 intron::GUS::NOS terminator (EXS1025)

[0216] EXS1025 is strongly expressed in the embryo. This expression starts between 5 days after pollination (DAP) and 10DAP. Expression is strongest in the scutellum and weaker in the embryo axis (plumule with leaves and internodes, primary root). Significant expression was also detected in the outermost layers of the kernel that include the aleuron layer. Expression is strongest at stages 15DAP to 25DAP and weaker at 30DAP. Weak expression was sometimes detected in the endosperm. No expression could be detected in any other organ including pollen.

6.5 Os.V-ATPase promoter::BPSI.1 intron::GUS::NOS terminator (pBPSMM369)

[0217] pBPSMM369 is strongly expressed in roots. This expression was detected in all tested stages. Significant expression was also detected in all parts of the kernels and in pollen. Weak expression was detected in the leaves at early developmental stages and at flowering. This expression was variable in strength and was in several plants at the detection limit. In general, expression was higher in homozygous T1 plants than in the heterozygous T0.

6.6 Zm.LDH promoter::BPSI.1 intron::GUS::NOS terminator (pBPSMM357)

[0218] pBPSMM357 shows weak activity in kernels. Expression in kernels was mainly located in and around the embryo. Very weak expression was also detected in roots.

6.7 Os.C8,7SI promoter::BPSI.1 intron::GUS::NOS terminator (pBPSMM366)

[0219] Os.C-8,7-sterol-isomerase promoter containing BPSI.1 (pBPSMM366) shows weak activity in roots and good expression in kernels.

6.8 Os.Lea promoter::BPSI.1 intron::GUS::NOS terminator (pBPSMM371)

[0220] Os.Lea promoter in combination with BPSI.1 (pBPSMM371) showed strong embryo-specific expression in kernels. Some expression could be detected in root tips but no expression was detected in any other tissue tested.

6.9 Zm.LDH promoter::BPSI.5 intron::GUS::NOS terminator (pBPSLM229)

[0221] pBPSLM229 shows weak expression in endosperm and aleuron layer, mainly at the top side of the kernel. No expression was detected in any other tissue tested.

Claims

1. A recombinant DNA expression construct comprising

a) at least one promoter sequence functioning in plants or plant cells, and
b) the intron consisting of the sequence described by SEQ ID NO: 2, and
c) at least one nucleic acid sequence,

wherein at least one of said promoter sequence and the intron sequence are functionally linked to at least one of said nucleic acid sequence and wherein said intron is heterologous to said nucleic acid sequence and/or to said promoter sequence.

2. The recombinant DNA expression construct of claim 1, wherein said functional equivalent comprises the functional elements of an intron and is **characterized by** a sequence

1. having at least 50 consecutive base pairs of an intron sequence described by SEQ ID NO: 2, or
2. having an identity of at least 80% over a sequence of at least 95 consecutive nucleic acid base pairs to a sequences described by SEQ ID NO: 2, or
3. hybridizing under high stringent conditions with a nucleic acid fragment of at least 50 consecutive base pairs of a nucleic acid molecule described by SEQ ID NO: 2.

3. The recombinant DNA expression construct of claim 1 to 2, further comprising one or more additional regulatory sequences functionally linked to said promoter.

4. The recombinant DNA expression construct of any of claim 1 to 3, wherein said promoter sequence functioning in plants or plant cells is selected from the group consisting of

a) the rice chloroplast protein 12 promoter as described by nucleotide 1 to 854 of SEQ ID NO: 113, or a sequence having at least 60% identity to said fragment, or a sequence hybridizing under stringent conditions to said fragment, or a sequence comprising at least 50 consecutive nucleotides of said fragment, and
b) the maize hydroxyproline-rich glycoprotein promoter as described by nucleotide 1 to 1184 of SEQ ID NO: 114, or a sequence having at least 60% identity to said fragment, or a sequence hybridizing under stringent conditions to said fragment, or a sequence comprising at least 50 consecutive nucleotides of said fragment, and
c) the p-caffeoyl-CoA 3-O-methyltransferase promoter as described by nucleotide 1 to 1034 of SEQ ID NO: 115, or a sequence having at least 60% identity to said fragment, or a sequence hybridizing under stringent conditions to said fragment, or a sequence comprising at least 50 consecutive nucleotides of said fragment, and
d) the maize Globulin-1 [ZmGIb1] promoter (W64A) as described by nucleotide 1 to 1440 of SEQ ID NO: 116, or a sequence having at least 60% identity to said fragment, or a sequence hybridizing under stringent conditions to said fragment, or a sequence comprising at least 50 consecutive nucleotides of said fragment, and
e) the putative Rice H+-transporting ATP synthase promoter as described by nucleotide 1 to 1589 of SEQ ID NO: 117, or a sequence having at least 60% identity to said fragment, or a sequence hybridizing under stringent conditions to said fragment, or a sequence comprising at least 50 consecutive nucleotides of said fragment, and
f) the putative rice C-8,7 sterol isomerase promoter as described by nucleotide 1 to 796 of SEQ ID NO: 118, or a sequence having at least 60% identity to said fragment, or a sequence hybridizing under stringent conditions to said fragment, or a sequence comprising at least 50 consecutive nucleotides of said fragment, and
g) the maize lactate dehydrogenase promoter as described by nucleotide 1 to 1062 of SEQ ID NO: 119, or a sequence having at least 60% identity to said fragment, or a sequence hybridizing under stringent conditions to said fragment, or a sequence comprising at least 50 consecutive nucleotides of said fragment, and
h) the rice Lea promoter as described by nucleotide 1 to 1386 of SEO ID NO: 121, or a sequence having at least 60% identity to said fragment, or a sequence hybridizing under stringent conditions to said fragment, or a sequence comprising at least 50 consecutive nucleotides of said fragment.

5. An expression vector comprising a recombinant expression construct of any of claim 1 to 4.

6. A transgenic cell or transgenic non-human organism comprising an expression vector as claimed in claim 5 or an expression construct of any of claim 1 to 4.

7. The cell or non-human organism of claim 6, selected from the group consisting of bacteria, fungi, yeasts and plants.

8. A cell culture, parts or propagation material derived from a transgenic cell organism of claim 6 to 7.

9. A method for providing an expression cassette for enhanced expression of a nucleic acid sequence in a plant or a plant cell, comprising the step of functionally linking at least one intron as described in claim 1 to 2 to said nucleic acid sequence.

10. A method for enhancing the expression of a nucleic acid sequence in a plant or a plant cell, comprising functionally linking at least one intron as described in claim 1 to 2 to said nucleic acid sequence.

11. The method as claimed in claim 9 to 10, wherein furthermore a promoter sequence functional in plants is linked to said nucleic acid sequence.

12. The method as claimed in claim 9 to 10, wherein the intron is linked to said nucleic acid sequence by insertion into the plant genomic DNA via homologous recombination.

13. The method of any of claim 9 to 12, wherein said plant or plant cell is a monocotyledonous plant or plant cell.

14. The use of a transgenic organism as claimed in claim 6 to 7 or of cell cultures, parts of transgenic propagation material derived there from as claimed in claim 8 for the production of foodstuffs, animal feeds, seeds, pharmaceuticals or fine chemicals.

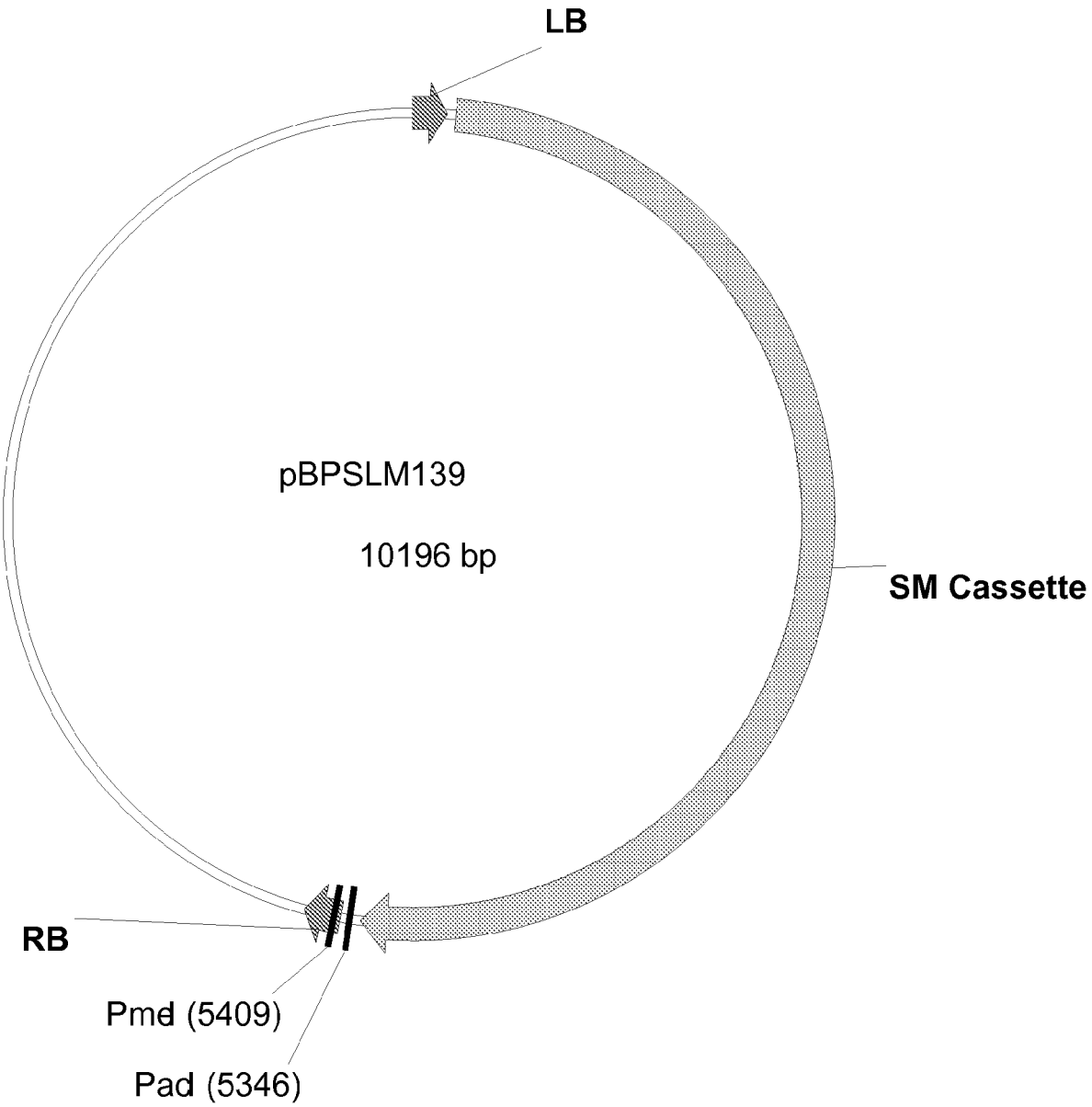

Fig. 1

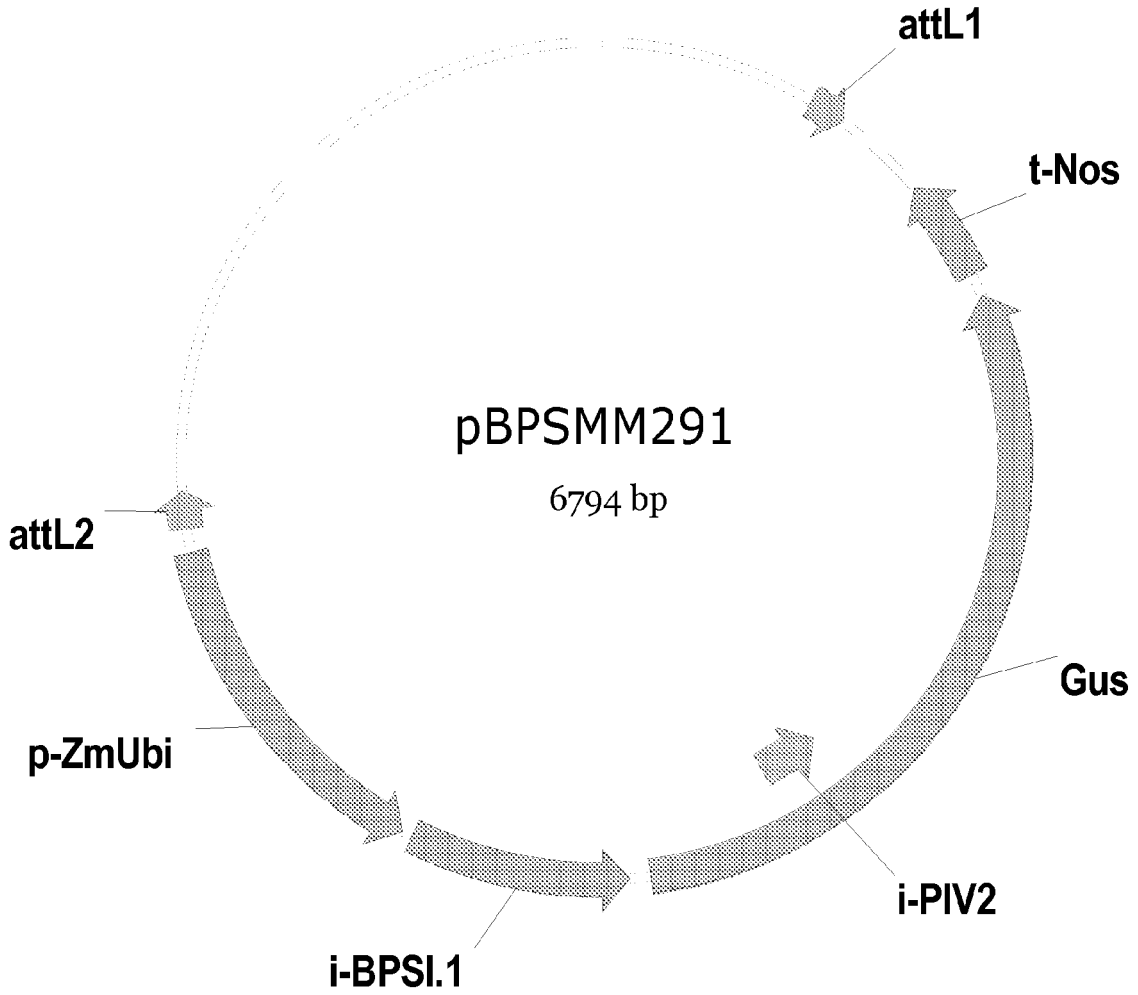

Fig. 2

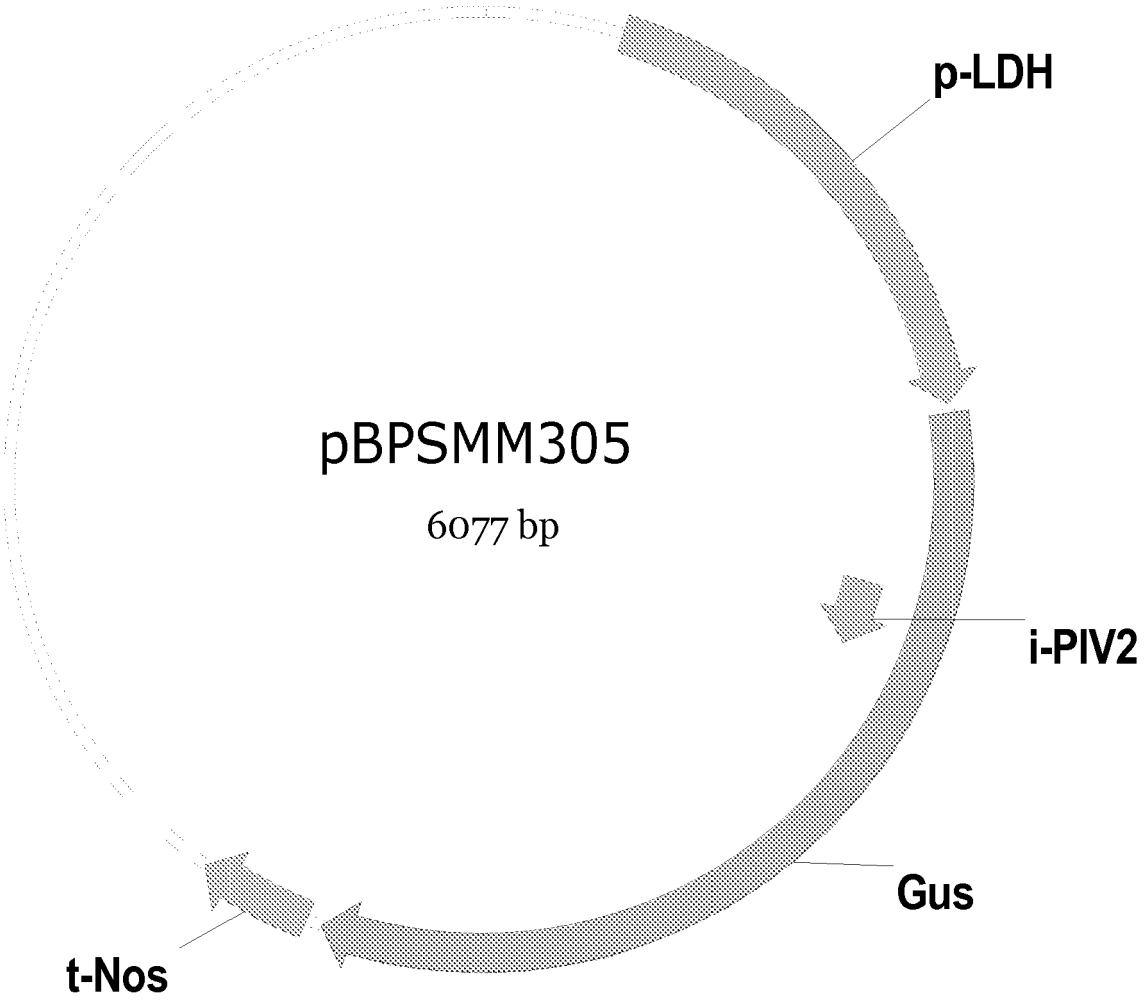

Fig. 3

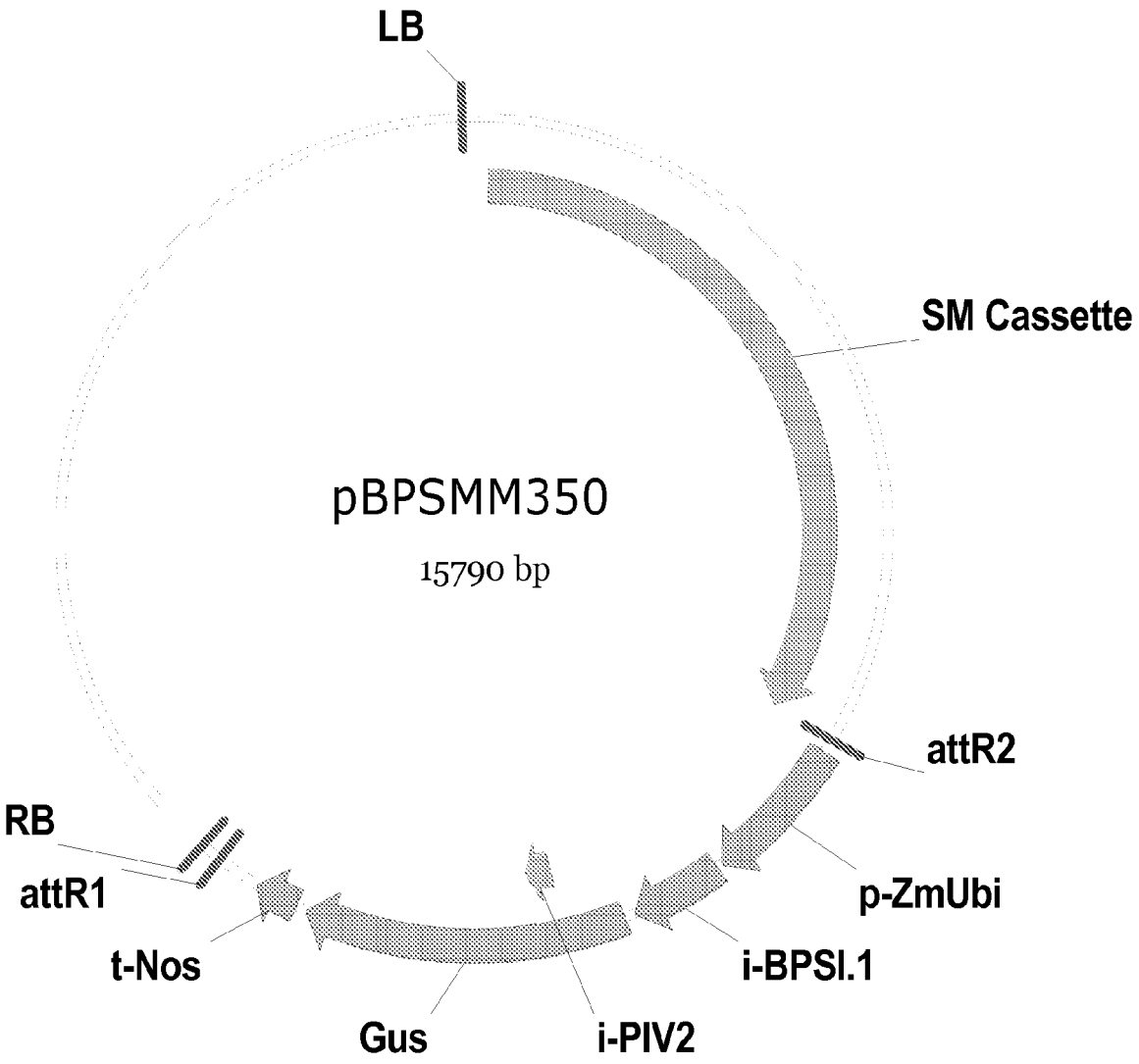
LB
SM Cassette
pBPSMM350
15790 bp
attR2
p-ZmUbi
i-BPSI.1
i-PIV2
Gus
t-Nos
attR1
RB

Fig. 4

```
#!/usr/local/bin/perl -w
# intron.pl

open(IN,$ARGV[0]) or die "can't find output";

 while (defined(my $file=<IN> )) {

#start of a single annotation

  if ($file=~/LOCUS.*?\s+(\d+)\sbp(.*)/) {
       my $length=$1;
       my $mol=1;
                    $mol=0 if $2 =~ /circular/;
                  my @cdslist=();
                  my @start=();
       my $order=0; # order=1: complementary coding.
                  my @title=();
                  my @title0=();
                  my @intron=();
                  my $id="";
                  my @terminator=();
                  my @promoter=();
                  my @utr5=();
                  my @utr3=();
                  my @origin=();
                  my $tab="";
                  my $organism="";
       while (defined(my $line=<IN> )) {
                  $line=$tab.$line;

                  if ($line =~ /^VERSION.*?\s+(GI:\d+)/) {
          $id=$1;

                  }elsif ($line =~ /^\s{2}ORGANISM\s+(.*)/){

                    if($1=~/Oryza sativa/i){
                      $organism="rice";
        }elsif($1=~/Zea mays/i) {
                      $organism="maize";
                    }elsif($1=~/Glycine max/i){
                      $organism="soybean";
                    }else {
                      $1=~/(\w+)/;
                                $organism=$1;
                    }
                  }elsif($line =~ /^\s{5}(CDS\s*)/){   #extract cds
                     my $test=$';
                     my $gene="N/A";
                                my $start=1;
                                my $product="N/A";
                                my $gi=$id;
                                my @cds=();
                                my @temp=();
                                if ($test =~ /complement/) {
                        $order=1 ;
                     }else {
                        $order = 0;
                     }
```

Fig. 5a

```
                while ( my $in=<IN>) {
                        if ($in =~ /\sV(.*)/) {
                          $test=$test;
                          if ($1=~/gene="(.*)"/) {
                            $gene=$1;
                          }elsif($1=~/note="(.*)"/) {
                            $product=$1;
                          }else {
                          last;
                          }
                        } else {
                          $test=$test.$in;
                        }
                      } #close while loop;

                        $test =~s/\w+\d+\.\d:\d+\.\.\d+//g;
                        $test =~ s/\D/ /g;
          $test =~ s/\s+/ /g;
          $test =~ s/^\s+//;
                my @sort;
                      if ($mol==0) {
                        @sort=split(/ /,$test);
                      } else {
        @sort=sort {$a <=> $b} split(/ /,$test);
                      }
# tag complement cds
                      if ($order==1) {
            @cds = ("complement",@sort);

                } elsif ($order==0) {
            @cds = @sort;
                } #close if loop;
#retreave notation if intron exist;
                if (scalar(@cds) >= 4) {

                  while (my $in=<IN>) {
                        $start=1;
          if ($in =~ /codon_start=(\d+)/) {
                        $start = $1;

                      }elsif ($in =~ /\/gene="(.*)"/){
                        $gene=$1;
                    }elsif ($in =~ /\/product=(.*)/){
                        $product=$1;
                        $product=~ tr/""//d;

                    }elsif ($in =~ /db_xref="(GI:.*?)"/) {
                        $gi = $1;

                        last ;
         } elsif ($in=~ /\/(pseudo)/) {
                        $product="pseudo";
                                last;
                      }             #close if loop
                    }  #close while loop;
                    push @start, $start;
                    push @cdslist, \@cds;
```

Fig. 5b

```
# retreave 5'utr if start codon > 1;
                                    my @tem=();
                                    for (my $i=1;$i<=($#cds-1)/2;$i++) {
                                      my $title1=">$organism|$gi|Intron_$i ";
                                      my                $title2="                $gene|$start|".($cds[2*$i-
1+$order]+1)."..".($cds[2*$i+$order]-1)."|$product\n";
                                      my @title=($title1,$title2);
                                      push @tem, \@title;
                                    } #close for loop
                                    push @title, \@tem;
                                    my    $title0=">$organism|$gi|5UTR_0       $gene|$start|".($cds[$order]-
1)."..".($cds[$order]+$start-2)."|$product\n";
                                    push @title0, $title0;

                          } #close if @cds>4 loop

        } elsif ($line =~ /^\s{5}terminator/) {

                          ($tab,my $note,my @term)=&getTerminator($line);

                                    push @terminator, $note;
                                    push @terminator, \@term;

                          }  elsif ($line =~ /^\s{5}promoter/) {

                             ($tab,my $note,my @prom)=&getTerminator($line);

                             push @promoter, $note;
                                    push @promoter, \@prom;

                          } elsif ($line =~ /^\s{5}5\DUTR/) {

                             ($tab,my $note,my @temp)=&getTerminator($line);

                        push @utr5,$note;
                                    push @utr5,\@temp;

                          } elsif ($line =~ /^\s{5}3\DUTR/) {

                             ($tab,my $note,my @temp)=&getTerminator($line);

                        push @utr3,$note;
                                    push @utr3,\@temp;

 #get sequence @origin
                          }
                         if ($line =~ /^(ORIGIN)/) {
                             $line="";
                                    while (my $code=<IN>) {

                                      if ($code =~ /\W/) {
                                 last;
                                  }else{
                                       $line=$line.$code;
                                      } #close if loop
                             } #close while loop
                                     # $line =~ s/\W/ /g;
                                    # print $line,"\n";
```

Fig. 5c

```
$line =~ tr/0-9//d;
        $line =~ tr/ //d;
        $line =~ tr/\n//d;
                        @origin = split(//,$line);


              for (my $i=0; $i<=$#cdslist;$i++) {

               if ($start[$i]>2) {

                 my @first=();
                 my $first;
                 if (${$cdslist[$i]}[0]  eq "complement") {
                   my @utr=@origin[$cdslist[$i][1]-1 .. ($cdslist[$i][1]+$start[$i]-2)];
                           print @utr,"\n";
                   $first=&complement(@utr);
                 } else {
                    @first=@origin[$cdslist[$i][0]-1 .. ($cdslist[$i][0]+$start[$i]-2)];
                           $first=join('',@first);
                 } #close if loop for complement

      print  $title0[$i],$first,"\n\n";
     }  #close if loop for $start>2;

               if (${$cdslist[$i]}[0] eq "complement") {
                 shift @{$cdslist[$i]};
                 for (my $j=1; $j<=($#{$cdslist[$i]}-1)/2;$j++) {
                   my @int=@origin[$cdslist[$i][2*$j-1] .. $cdslist[$i][2*$j]-2];
                           my $int1=&complement(@int);
                           print      $title[$i][$j-1][0],scalar(@int),$title[$i][$j-1][1],    $int1,"\n\n"    if
$#int<5000;
                 } #close 2nd for loop for complement

               } else {
                 for (my $j=1; $j<=($#{$cdslist[$i]}-1)/2;$j++) {
                           my @int=@origin[$cdslist[$i][2*$j-1] .. $cdslist[$i][2*$j]-2];

                           if ($mol==0 && $cdslist[$i][2*$j-1] > $cdslist[$i][2*$j]) {
                             @int=(@origin[$cdslist[$i][2*$j-1]      ..      $#origin],      @origin[0      ..
$cdslist[$i][2*$j]-2]);
                  }
                           my $int1=join('',@int);
                  print  $title[$i][$j-1][0],scalar(@int),$title[$i][$j-1][1], $int1,"\n\n" if $#int < 5000;
                          }#close 2nd for loop
      } #close else loop
     } #close 1st for loop

               my $title1=">$organism|$id|terminator";
               &getSequence(\@terminator,\@origin,$title1);

                $title1=">$organism|$id|promoter";
               &getSequence(\@promoter,\@origin,$title1);

                $title1=">$organism|$id|5utr";
               &getSequence(\@utr5,\@origin,$title1);

                $title1=">$organism|$id|3utr";
               &getSequence(\@utr3,\@origin,$title1);
```

Fig. 5d

```
                    last;
                   } else {
                    $tab="";
      } #close if $line loop
     } #close while $line loop
     next;

   } #close if $file loop

  } #close while $file loop
close IN;

#retreave complement sequnce
sub complement{
  my @code=@_;
  my @complemnt=();
  for (my $i=0;$i<=$#code;$i++) {
    if ($code[$#code-$i] eq "t") {
       $complement[$i]= "a";
    } elsif ($code[$#code-$i] eq "a") {
       $complement[$i]= "t";
    } elsif ($code[$#code-$i] eq "c") {
       $complement[$i]= "g";
    } elsif ($code[$#code-$i] eq "g") {
                $complement[$i]= "c";
    } else {
       $complement[$i]=$code[$#code-$i];
    }#close if loop
  } #close for loop
  my $comp=join('',@complement);
  @complement=();
  return $comp;
} #close sub

#get sequence reference for feature keys
sub getTerminator {
 my $line=$_[0];
 my $order=0;
 if ($line=~/complement/) {
    $order=1;
 } else {
 } #close if loop
 $line =~ s/\d'UTR//;
 $line =~ s/\D/ /g;
 $line =~ s/\s+/ /g;
 $line =~ s/^\s//;
 my @term=split(' ',$line);
   @term=("c",@term) if $order==1;
 my $in;
 read(IN,$in,6);
 my $note =" \n";
```

Fig. 5e

```
if ($in!~/\w/) {
   $note=<IN>;
   $note=~s/\s+\//;
   $note=~s/note=//;
   $note=~ tr/""//d;
 } #close if loop
 return ($in,$note,@term);
} #close sub


#retreave sequence information for feature keys
sub getSequence {
my @array=@{$_[0]};
my @code=@{$_[1]};
my $id=$_[2];

 for (my $i=0; $i<($#array+1)/2;$i++) {
     my $note=$array[2*$i];
     my @term=@{$array[2*$i+1]};
   if ($term[0] eq "c") {
     shift @term;
     for (my $j=0; $j<=($#term-1)/2;$j++) {
       my @comp=@code[($term[2*$j]-1) .. ($term[2*$j+1]-1)];
       my $int1=&complement(@comp);
       my $title=$id."_".($i+1)." ".scalar(@comp)."  $term[2*$j]..$term[2*$j+1]|$note";
       print $title, $int1,"\n\n";
     } #close 2nd for loop
   } else {
     for (my $j=0; $j<($#term+1)/2;$j++) {
                 my @int=@code[($term[2*$j]-1) .. ($term[2*$j+1]-1)];
       my $int1=join(",@int);
       my $title=$id."_".($i+1)." ".scalar(@int)."  $term[2*$j]..$term[2*$j+1]|$note";
                 print $title, $int1,"\n\n";
     } #close 2nd for loop
   } #close if loop
  } #close 1st for loop
} #close sub
```

Fig. 5f

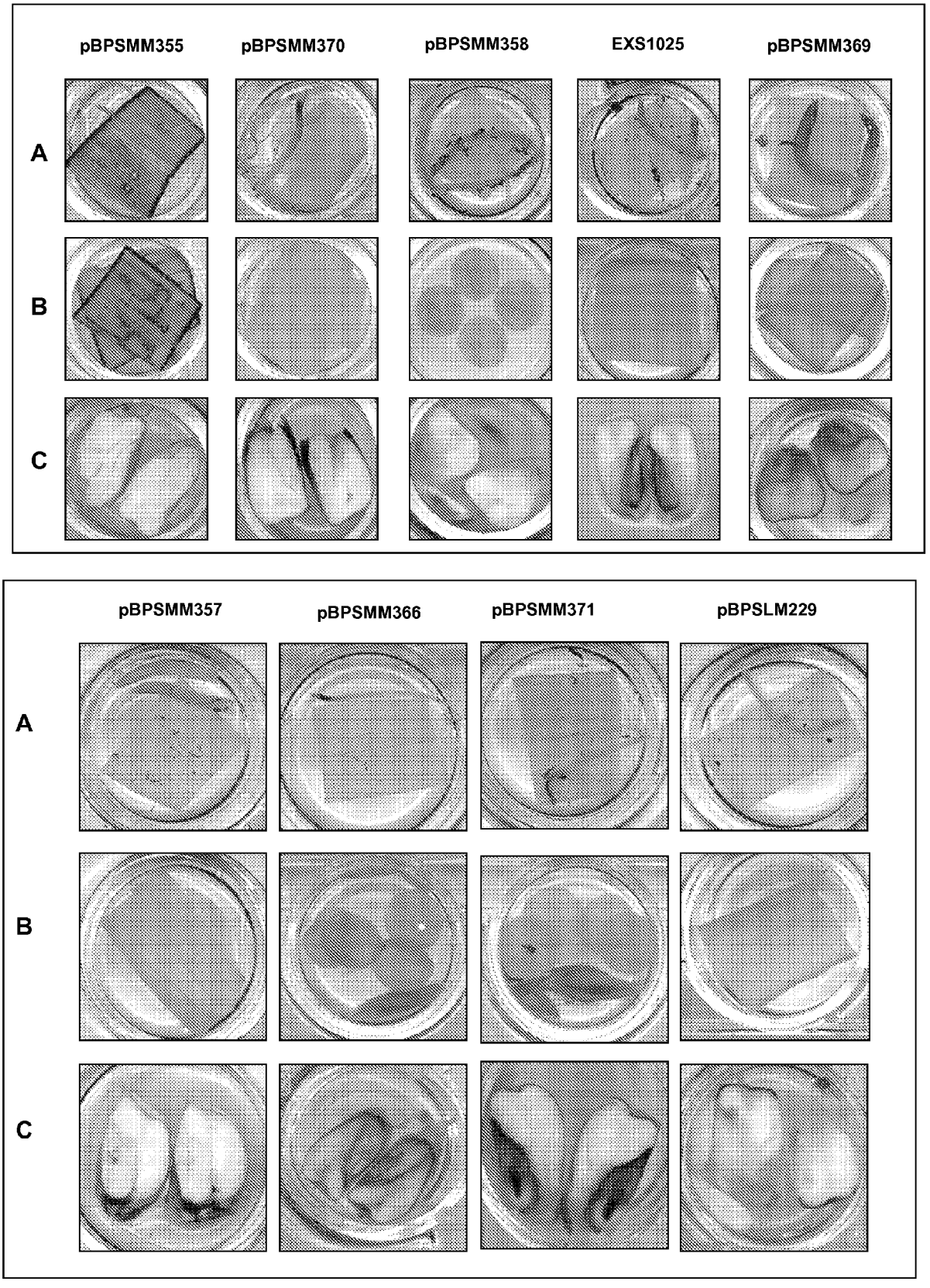
pBPSMM355
pBPSMM370
pBPSMM358
EXS1025
pBPSMM369
A
B
C
pBPSMM357
pBPSMM366
pBPSMM371
pBPSLM229
A
B
C

Fig. 6

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 5584807 A **[0040]**
- WO 9932619 A **[0061]**
- WO 9953050 A **[0061]**
- WO 0068374 A **[0061]**
- WO 0044914 A **[0061]**
- WO 0044895 A **[0061]**
- WO 0049035 A **[0061]**
- WO 0063364 A **[0061]**
- US 5352605 A **[0079]**
- WO 8402913 A **[0079]**
- US 4962028 A **[0079]**
- WO 9519443 A **[0079]**
- EP 335528 A **[0079]**
- WO 9321334 A **[0079]**
- US 5608152 A **[0079]**
- WO 9845461 A **[0079]**
- US 5504200 A **[0079]**
- WO 9113980 A **[0079]**
- WO 0026388 A **[0079]**
- WO 9515389 A **[0079]**
- WO 9523230 A **[0079]**
- WO 9916890 A **[0079]**
- US 5677474 A **[0079]**
- US 5530149 A **[0079]**
- EP 571741 A **[0079]**
- JP 6062870 A **[0079]**
- WO 9808962 A **[0079]**
- US 5689040 A **[0079]**
- EP 781849 A **[0079]**
- DE 19644478 A **[0079]**
- EP 0249676 A **[0079]**
- US 5187267 A **[0079]**
- WO 9612814 A **[0079]**
- EP 0375091 A **[0079]**
- WO 9516783 A **[0079]**
- WO 9706250 A **[0079]**
- WO 9946394 A **[0079]**
- WO 0114572 A **[0079]**
- WO 9514098 A **[0079]**
- WO 9412015 A **[0079]**
- EP 388186 A **[0079]**
- WO 9706268 A **[0079]**
- US 5290924 A **[0079]**
- US 5565350 A **[0081]**
- WO 0015815 A **[0081]**
- US US5667972 A **[0108]**
- US 5492806 A **[0108]**
- US 5695940 A **[0108]**
- US 5972619 A **[0108]**
- US 6018041 A **[0108]**
- US 6451996 B **[0108]**
- US 6309824 B **[0108]**
- WO 9845456 A **[0118] [0119]**
- EP 0807836 A **[0119]**
- WO 03060133 A **[0119]**
- EP 0601092 A **[0121]**
- WO 9741228 A **[0122]**
- WO 2001007596 A **[0125]**
- US 2002138870 A **[0125]**
- US 5625047 A **[0125]**
- WO 0000512 A **[0127]**
- WO 9712983 A **[0127]**
- WO 9811240 A **[0127]**
- WO 9713843 A **[0127]**
- WO 9826045 A **[0127]**
- WO 9905902 A **[0127]**
- WO 9906580 A **[0127]**
- WO 9904013 A **[0127]**
- WO 9742326 A **[0127]**
- WO 9750561 A **[0127]**
- EP 120516 A **[0138]**
- US 4940838 A **[0138]**
- WO 0200900 A **[0138]**
- WO 9102071 A **[0140]**
- WO 9318168 A **[0140]**
- WO 0063398 A **[0140]**
- WO 9400583 A **[0140]**
- WO 9748814 A **[0140]**
- US 5501967 A **[0147]**
- US 6110736 A **[0148]**
- WO 9920780 A **[0149] [0151]**
- WO 9708331 A **[0150] [0152]**
- WO 0168882 A **[0150]**
- WO 0138504 A **[0150]**
- US 5527695 A **[0151]**
- WO 0026192 A **[0152]**
- WO 0029384 A **[0152]**
- WO 0032579 A **[0152]**
- WO 0064878 A **[0152]**
- WO 0068206 A **[0152]**
- WO 0067734 A **[0152]**
- WO 0123386 A **[0152]**
- WO 0123390 A **[0152]**
- EP 0424047 A1 **[0210]**
- EP 0397687 A1 **[0210]**
- US 5376543 A **[0210]**
- US 5169770 A **[0210]**
- US 5591616 A **[0210]**

Non-patent literature cited in the description

- **Mascarenhas et al.** *Plant Mol. Biol.,* 1990, vol. 15, 913-920 **[0003]**
- **Jeon et al.** *Plant Physiol.,* 2000, vol. 123, 1005-1014 **[0003]**
- **Callis et al.** *Genes Dev.,* 1987, vol. 1, 1183-1200 **[0003]**
- **Vasil et al.** *Plant Physiol,* 1989, vol. 91, 1575-1579 **[0003]**
- **Christiansen et al.** *Plant Mol. Biol.,* 1992, vol. 18, 675-689 **[0003]**
- **McElroy et al.** *Plant Cell,* 1990, vol. 2, 163-171 **[0003]**
- **Xu et al.** *Plant Physiol,* 1994, vol. 106, 459-467 **[0003]**
- **Dean et al.** *Plant Cell,* 1989, vol. 1, 201-208 **[0003]**
- **Leon et al.** *Plant Phyisiol.,* 1991, vol. 95, 968-972 **[0003]**
- **Norris et al.** *Plant Mol Biol,* 1993, vol. 21, 895-906 **[0003]**
- **Rose ; Last.** *Plant J,* 1997, vol. 11, 455-464 **[0003]**
- **Mascarenhas et al.** *Plant Mol Biol,* 1990, vol. 15, 913-920 **[0003] [0004]**
- **Clancy ; Hannah.** *Plant Physiol,* 2002, vol. 130, 918-929 **[0003]**
- **Rose ; Beliakoff.** *Plant Physiol,* 2000, vol. 122, 535-542 **[0003]**
- **Chee et al.** *Gene,* 1986, vol. 41, 47-57 **[0004]**
- **Kuhlemeier et al.** *Mol Gen Genet,* 1988, vol. 212, 405-411 **[0004]**
- **Sinibaldi ; Mettler.** Progress in Nucleic Acid Research and Molecular Biology. Academic Press, 1992, vol. 42, 229-257 **[0004]**
- **Vancanneyt et al.** *Mol Gen Gent,* 1990, vol. 220, 245-250 **[0004]**
- **Rhodes.** *Methods Mol Biol,* 1995, vol. 55, 121-131 **[0049]**
- **Jefferson et al.** GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.,* 20 December 1987, vol. 6 (13), 3901-3907 **[0049]**
- **Wigler.** *Proc Natl Acad Sci,* 1980, vol. 77, 3567-3570 **[0049]**
- **Colbere-Garapin.** *J. Mol. Biol.,* 1981, vol. 150, 1-14 **[0049]**
- **Maniatis T ; Fritsch EF ; Sambrook J.** Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, 1989 **[0057]**
- **Silhavy TJ ; Berman ML ; Enquist LW.** Experiments with Gene Fusions. Cold Spring Harbor Laboratory, 1984 **[0057]**
- **Ausubel FM et al.** Current Protocols in Molecular Biology. Greene Publishing Assoc. and Wiley Interscience, 1987 **[0057]**
- **Gelvin et al.** Plant Molecular Biology Manual. 1990 **[0057]**
- **Goring.** *Proc. Natl Acad. Sci. USA,* 1991, vol. 88, 1770-1774 **[0061]**
- **Smith.** *Mol. Gen. Genet.,* 1990, vol. 224, 447-481 **[0061]**
- **Napoli.** *Plant Cell,* 1990, vol. 2, 279-289 **[0061]**
- **Van der Krol.** *Plant Cell,* 1990, vol. 2, 291-299 **[0061]**
- **Coombs.** Dictionary of Biotechnology. Stockton Press, 1994 **[0064]**
- **Anderson ; Young.** Quantitative Filter Hybridization. *Nucleic Acid Hybridization,* 1985 **[0064]**
- **Sambrook.** Molecular Cloning; A Laboratory Manual. Cold Spring Harbor Laboratory Press, 1989 **[0064]**
- Current Protocols in Molecular Biology. John Wiley & Sons, 1989, 6.3.1-6.3.6 **[0064]**
- **Benfey et al.** *EMBO J.,* 1989, vol. 8, 2195-2202 **[0079]**
- **Franck et al.** *Cell,* 1980, vol. 21, 285-294 **[0079]**
- **Sanger et al.** *Plant. Mol. Biol.,* 1990, vol. 14, 433-443 **[0079]**
- **Ward et al.** *Plant. Mol. Biol.,* 1993, vol. 22, 361-6 **[0079]**
- **Leisner.** *Proc Natl Acad Sci USA,* 1988, vol. 85 (5), 2553-2557 **[0079]**
- **Comai.** *Plant Mol Biol,* 1990, vol. 15 (3), 373-381 **[0079]**
- **Schenk.** *Plant Mol Biol,* 1999, vol. 39 (6), 1221-1230 **[0079]**
- **Jain et al.** *Crop Science,* 1999, vol. 39 (6), 1696-1701 **[0079]**
- **Shaw et al.** *Nucleic Acids Res.,* 1984, vol. 12 (20), 7831-7846 **[0079]**
- **Gatz.** *Annu. Rev. Plant Physiol. Plant Mol. Biol.,* 1997, vol. 48, 89-108 **[0079]**
- **Gatz et al.** *Plant J.,* 1992, vol. 2, 397-404 **[0079]**
- **Huang et al.** *Cell,* 1981, vol. 27, 245 **[0079]**
- **Chua et al.** *Science,* 1989, vol. 244, 174-181 **[0079]**
- **Bäumlein et al.** *Mol Gen Genet,* 1991, vol. 225 (3), 459-67 **[0079]**
- **Bäumlein et al.** *Plant Journal,* 1992, vol. 2 (2), 233-9 **[0079]**
- **Stockhaus et al.** *EMBO J.,* 1989, vol. 8, 2445 **[0079]**
- **Ward et al.** *Plant. Mol. Biol.,* 1993, vol. 22, 361-366 **[0079]**
- **Xu et al.** *Plant Cell,* 1994, vol. 6 (8), 1077-1085 **[0079]**
- **Tran et al.** *Plant Cell,* 2004, vol. 16 (9), 2481-2498 **[0079]**
- **Maniatis.** Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, 1989 **[0081]**
- **Silhavy.** Experiments with Gene Fusions. Cold Spring Harbor Laboratory, 1984 **[0081]**
- **Ausubel.** Current Protocols in Molecular Biology. Greene Publishing Assoc. and Wiley Interscience, 1987 **[0081]**
- **Mozo et al.** *Mol. Gen. Genet.,* 1998, vol. 258, 562-570 **[0084]**
- **Choi et al.** *Weeds World,* 1995, vol. 2, 17-20 **[0084]**

- **Lui et al.** *Proc. Natl. Acad. Sci. USA,* 1999, vol. 96, 6535-6540 **[0084]**
- The Arabidopsis Genome initiative. *Nature,* 1999, vol. 402, 761-777 **[0084]**
- The Arabidopsis Genome initiative. *Nature,* 2000, vol. 408, 796-826 **[0084]**
- **Milosavljevic, A. et al.** *Genome Res.,* 1996, vol. 6, 132-141 **[0104]**
- **Shoemaker, D. et al.** *Nature Genet.,* 1996, vol. 14, 450-456 **[0104]**
- **Sikela,J.M. ; Auffray,C.** *Nature Genet.,* 1993, vol. 3, 189-191 **[0104]**
- **Meier-Ewert S. et al.** *Nucleic Acids Research,* 1998, vol. 26 (9), 2216-2223 **[0104]**
- **Strezoska, Z. et al.** *Proc. Natl. Acad. Sci. USA,* 1991, vol. 88, 10089-10093 **[0104]**
- **Velculescu, V. E. et al.** *Science,* 1995, vol. 270, 484-487 **[0104]**
- **Bray.** *Trends Plant Sci.,* 1997, vol. 2, 48-54 **[0117]**
- **Busk et al.** *Plant Mol. Biol.,* 1998, vol. 37, 425-435 **[0117]**
- **Shinozaki ; Yamaguchi-Shinozaki.** *Curr. Opin. Plant Biol.,* 2000, vol. 3, 217-223 **[0117]**
- **Guiltinan et al.** *Science,* 1990, vol. 250, 267-271 **[0117]**
- **Mundy et al.** *Proc. Natl. Acad. Sci. USA,* 1990, vol. 87, 406-410 **[0117]**
- **Yamaguchi-Shinozaki ; Shinozaki.** *Plant Cell,* 1994, vol. 6, 251-264 **[0117]**
- **Wang et al.** *Plant Mol. Biol.,* 1995, vol. 28, 605-617 **[0117]**
- **Iwasaki et al.** *Plant Physiol.,* 1997, vol. 115, 1287 **[0117]**
- **Amin et al.** *Mol Cell Biol,* 1988, vol. 8, 3761-3769 **[0117]**
- **Eulgem et al.** *Trends Plant Sci,* 2000, vol. 5, 199-206 **[0117]**
- **Robatzek S et al.** *Plant J,* 2001, vol. 28, 123-133 **[0117]**
- **Rushton PJ et al.** *Curr Opin Plant Biol,* 1998, vol. 1, 311-315 **[0117]**
- **Gallie et al.** *Nucl Acids Res,* 1987, vol. 15, 8693-8711 **[0117]**
- **Rouster J et al.** *Plant J,* 1998, vol. 15, 435-440 **[0117]**
- **Gielen et al.** *EMBO J,* 1984, vol. 3, 835-46 **[0117]**
- **McCormick et al.** *Plant Cell Reports,* 1986, vol. 5, 81-84 **[0118]**
- **de Block et al.** *EMBO J,* 1987, vol. 6, 2513-2518 **[0119]**
- **Randez-Gil et al.** *Yeast,* 1995, vol. 11, 1233-1240 **[0119]**
- **Koprek T et al.** *Plant J,* 1999, vol. 19 (6), 719-726 **[0120]**
- **Gleave AP et al.** *Plant Mol Biol.,* 1999, vol. 40 (2), 223-35 **[0120]**
- **Perera RJ et al.** *Plant Mol. Biol,* 1993, vol. 23 (4), 793-799 **[0120]**
- **Stougaard J.** *Plant J,* 1993, vol. 3, 755-761 **[0120]**
- **Koprek et al.** *Plant J,* 1999, vol. 16, 719-726 **[0120]**
- **Naested H.** *Plant J,* 1999, vol. 18, 571-576 **[0120]**
- **Sundaresan V.** *Genes & Development,* 1995, vol. 9, 1797-1810 **[0120]**
- **Schlaman HRM ; Hooykaas PJJ.** *Plant J,* 1997, vol. 11, 1377-1385 **[0120]**
- **Fedoroff NV ; Smith DL.** *Plant J,* 1993, vol. 3, 273-289 **[0120]**
- **Schenborn E ; Groskreutz D.** *Mol Biotechnol.,* 1999, vol. 13 (1), 29-44 **[0122]**
- **Sheen et al.** *Plant Journal,* 1995, vol. 8 (5), 777-784 **[0122]**
- **Haseloff et al.** *Proc Natl Acad Sci USA,* 1997, vol. 94 (6), 2122-2127 **[0122]**
- **Reichel et al.** *Proc Natl Acad Sci USA,* 1996, vol. 93 (12), 5888-5893 **[0122]**
- **Tian et al.** *Plant Cell Rep,* 1997, vol. 16, 267-271 **[0122]**
- **Chui WL et al.** *Curr Biol,* 1996, vol. 6, 325-330 **[0122]**
- **Leffel SM et al.** *Biotechniques,* 1997, vol. 23 (5), 912-8 **[0122]**
- **Ow et al.** *Science,* 1986, vol. 234, 856-859 **[0122]**
- **Millar et al.** *Plant Mol Biol Rep,* 1992, vol. 10, 324-414 **[0122]**
- **Prasher et al.** *Biochem Biophys Res Commun,* 1985, vol. 126 (3), 1259-1268 **[0122]**
- **Dellaporta et al.** Chromosome Structure and Function: Impact of New Concepts. *18th Stadler Genetics Symposium,* 1988, vol. 11, 263-282 **[0122]**
- **Jefferson et al.** *EMBO J.,* 1987, vol. 6, 3901-3907 **[0122]**
- **Sambrook et al.** Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, 1989 **[0123]**
- **Ullah et al.** *Plant Cell,* 2003, vol. 15, 393-409 **[0125]**
- **Kim ; Ceck.** *Proc. Natl. Acad. Sci. USA,* 1987, vol. 84, 8788-8792 **[0125]**
- **Gerlach et al.** *Nature,* 1987, vol. 328, 802-805 **[0125]**
- **Forster ; Symons.** *Cell,* 1987, vol. 49, 211-220 **[0125]**
- **Symons.** *Annu. Rev. Biochem.,* 1992, vol. 61, 641-671 **[0125]**
- **Prody et al.** *Science,* 1986, vol. 231, 1577-1580 **[0125]**
- **Yan et al.** *Proc. Natl. Acad. Sci. USA,* 1992, vol. 87, 4144-4148 **[0125]**
- **Berzal-Herranz et al.** *Genes and Devel.,* 1992, vol. 98, 1207-1210 **[0125]**
- **Haseloff ; Gerlach.** *Nature,* 1988, vol. 224, 585-591 **[0125]**
- **Beerli.** *Proc Natl Acad Sci USA,* 2000, vol. 97 (4), 1495-500 **[0126]**
- **Sauer B.** *Methods,* 1998, vol. 14 (4), 381-92 **[0128]**
- **Hsieh, H.M. et al.** RNA expression patterns of a type 2 metallothioneine-like gene from rice. *Plant Mol. Biol.,* 1996, vol. 32 (3), 525-529 **[0129]**

- **Altschul, Stephen F. et al.** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.,* 1997, vol. 25, 3389-3402 **[0129]**
- **Kado.** *Crit Rev Plant Sci,* 1991, vol. 10, 1 **[0138]**
- **Zupan.** *Plant J,* 2000, vol. 23 (1), 11-28 **[0138]**
- **Miki.** Procedures for Introducing Foreign DNA into Plants. *METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY,* 1993, 67-88 **[0138]**
- **Gruber.** Vectors for Plant Transformation. *METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY,* 1993, 89-119 **[0138]**
- **Moloney.** *Plant Cell Reports,* 1989, vol. 8, 238-242 **[0138] [0210]**
- **Hoekema.** The Binary Plant Vector System. Offsetdrukkerij Kanters B.V, 1985 **[0138]**
- **Fraley.** *CRC Crit. Rev. Plant. Sci.,* 1985, vol. 4, 1-45 **[0138]**
- **An.** *EMBO J.,* 1985, vol. 4, 277-287 **[0138]**
- **Holsters.** *Mol Gen Genet,* 1978, vol. 163, 181-187 **[0138]**
- **Hoekema.** *Nature,* 1985, vol. 303, 179-181 **[0138]**
- **Bevan.** *Nucl Acid Res,* 1984, vol. 12, 8711-8720 **[0138]**
- **Watson.** *EMBO J,* 1985, vol. 4 (2), 277-284 **[0138]**
- **Hajdukiewicz.** *Plant Mol Biol,* 1994, vol. 25, 989-994 **[0138]**
- **Hood.** *J Bacteriol,* 1986, vol. 168, 1291-1301 **[0138]**
- **Li.** *Plant Mol Biol,* 1992, vol. 20, 1037-1048 **[0138]**
- **Hoekema.** *Nature,* 1983, vol. 303, 179-181 **[0138]**
- **Koncz.** *Mol Gen Genet,* 1986, vol. 204, 383-396 **[0138]**
- **Deblaere.** *Nucl Acids Res,* 1985, vol. 13, 4777-4788 **[0138]**
- **Van Larebeke.** *Nature,* 1974, vol. 252, 169-170 **[0138]**
- **Watson.** *J. Bacteriol,* 1975, vol. 123, 255-264 **[0138]**
- **Klapwijk.** *J. Bacteriol.,* 1980, vol. 141, 128-136 **[0138]**
- **Jarchow.** *Proc. Natl. Acad. Sci. USA,* 1991, vol. 88, 10426-10430 **[0138]**
- **Hansen.** *Proc. Natl. Acad. Sci. USA,* 1994, vol. 91, 7603-7607 **[0138]**
- **Chen.** *J. Bacteriol.,* 1991, vol. 173, 1139-1144 **[0138]**
- **Scheeren-Groot.** *J. Bacteriol,* 1994, vol. 176, 6418-6426 **[0138]**
- **Mozo.** *Plant Mol. Biol.,* 1991, vol. 16, 917-918 **[0138]**
- **Keown.** *Methods in Enzymology,* 1990, vol. 185, 527-537 **[0140]**
- **Nan.** Biotechnology in Agriculture and Forestry. Springer-Verlag, 1995, vol. 34, 145-155 **[0140]**
- **Griesbach.** *Hort Science,* 1992, vol. 27, 620 **[0140]**
- **Fraley.** *Proc. Natl. Acad. Sci. USA,* 1982, vol. 79, 1859-1863 **[0140]**
- **Krens.** *Nature,* 1982, vol. 296, 72-74 **[0140]**
- **Shimamoto.** *Nature,* 1989, vol. 338, 274-276 **[0140]**
- **Datta.** *Bio/Technology,* 1990, vol. 8, 736-740 **[0140]**
- **Paszkowski.** *EMBO J,* 1984, vol. 3, 2717-2722 **[0140]**
- **Shillito.** *Bio/Technology,* 1985, vol. 3, 1099-1103 **[0140]**
- **Fromm.** *Nature,* 1986, vol. 319, 791-793 **[0140]**
- **Hayakawa.** *Proc Natl Acad Sci USA,* 1992, vol. 89, 9865-9869 **[0140]**
- **Sanford.** *Physiologia Plantarium,* 1990, vol. 79, 206-209 **[0140]**
- **Fromm.** *Bio/Technology,* 1990, vol. 8, 833-839 **[0140]**
- **Christou.** *Plant Physiol,* 1988, vol. 87, 671-674 **[0140]**
- **Sautter.** *Bio/Technology,* 1991, vol. 9, 1080-1085 **[0140]**
- **Christou.** *Bio/Technology,* 1991, vol. 9, 957-962 **[0140]**
- **Gordon-Kamm.** *Plant Cell,* 1990, vol. 2, 603-618 **[0140]**
- **Vasil.** *Bio/Technology,* 1992, vol. 10, 667-674 **[0140]**
- *Bio/Technology,* 1993, vol. 11, 1153-1158 **[0140]**
- **Wan.** *Plant Physiol.,* 1994, vol. 104, 3748 **[0140]**
- **Somers.** *Bio/Technology,* 1992, vol. 10, 1589-1594 **[0140]**
- Techniques for Gene Transfer. **Jenes.** Transgenic Plants, Vol. 1, Engineering and Utilization. Academic Press, 1983, vol. 1, 128-143 **[0140]**
- **Potrykus.** *Ann Rev Plant Physiol Plant Mol Biol,* 1991, vol. 42, 205-225 **[0140]**
- **Shahla.** *Plant Mol. Biol.,* 1987, vol. 8, 291-298 **[0140]**
- **Bidney.** *Plant Molec. Biol.,* 1992, vol. 18, 301-313 **[0140]**
- **Luo.** *Plant Mol. Biol. Rep.,* 1988, vol. 6, 165-174 **[0140]**
- **Du.** *Genet Manip Plants,* 1989, vol. 5, 8-12 **[0140]**
- **Töpfer.** *Plant Cell,* 1989, vol. 1, 133-139 **[0140]**
- **McCormick.** *Plant Cell Reports,* 1986, vol. 5, 81-84 **[0141]**
- **Thomas KR ; Capecchi MR.** *Cell,* 1987, vol. 51, 503 **[0147]**
- **Strepp et al.** *Proc Natl Acad Sci USA,* 1998, vol. 95 (8), 4368-4373 **[0147]**
- **Rong ; Golic.** *Science,* 16 June 2000, vol. 288 (5473), 2013-8 **[0149]**
- **Schlaman ; Hooykaas.** *Plant J,* 1997, vol. 11, 1377-1385 **[0150]**
- **Thykjaer et al.** *Plant Mol Biol.,* November 1997, vol. 35 (4), 523-30 **[0150]**
- **Risseeuw et al.** *Plant J.,* April 1997, vol. 11 (4), 717-28 **[0150]**
- **Gallego et al.** *Plant Mol Biol.,* January 1999, vol. 39 (1), 83-93 **[0150]**
- **Terada et al.** *Nat Biotechnol.,* 09 September 2002, vol. 20 (10), 1030-4 **[0150]**
- **Puchta H et al.** *Plant J,* 1995, vol. 7, 203-210 **[0150]**
- **Puchta H et al.** *Plant J,* 1995, vol. 7 (2), 203-210 **[0150]**

- **Lebel EG et al.** *Proc Natl Acad Sci USA,* 1993, vol. 90 (2), 422-426 **[0150]**
- **Siebert R ; Puchta H.** *Plant Cell,* 2002, vol. 14 (5), 1121-1131 **[0150]**
- **Reiss B et al.** *Proc Natl Acad Sci USA,* 1996, vol. 93 (7), 3094-3098 **[0152]**
- **Reiss B et al.** *Proc Natl Acad Sci USA,* 2000, vol. 97 (7), 3358-3363 **[0152]**
- **Puchta H et al.** *Plant J.,* 1995, vol. 7, 203-210 **[0152]**
- **Hood.** *Curr Opin Biotechnol.,* 1999, vol. 10 (4), 382-6 **[0154]**
- **Ma.** *Curr Top Microbiol. Immunol.,* 1999, vol. 236, 275-92 **[0154]**
- **Russel.** *Current Topics in Microbiology and Immunology,* 1999, vol. 240, 119-138 **[0154]**
- **Cramer et al.** *Current Topics in Microbiology and Immunology,* 1999, vol. 240, 95-118 **[0154]**
- **Gavilondo.** *Biotechniques,* 2000, vol. 29 (1), 128-138 **[0154]**
- **Holliger.** *Cancer & Metastasis Reviews,* 1999, vol. 18 (4), 411-419 **[0154]**
- **Gubler U ; Hoffman BJ.** A simple and very efficient method for generating cDNA libraries. *Gene,* 1983, vol. 25 (2-3), 263-269 **[0158]**
- **Jung-Hwa Oh et al.** An improved method for constructing a full-length enriched cDNA library using small amounts of total RNA as a starting material. *EXPERIMENTAL and MOLECULAR MEDICINE,* 2003, vol. 35 (6), 586-590 **[0158]**
- **Lanfranchi et al.** Identification of 4370 expressed sequence tags from a 3'-end-specific cDNA library of human skeletal muscle by DNA sequencing and filter hybridization. *Genome Res.,* 1996, vol. 6 (1), 35-42 **[0158]**
- cDNA Library Protocols. **Cowell ; Austin.** Methods in Molecular Biology. Humana Press, Scientific and medical publishers, October 1996, vol. 69 **[0158]**
- Generation of cDNA Libraries, Methods and Protocols. **Ying, Shao-Yao.** Methods in Molecular Biology. Humana Press, Scientific and medical publishers, February 2003, vol. 221 **[0158]**
- **Gelvin ; Schilperoort.** Plant Molecular Biology Manual. Kluwer Academic Publ, 1995 **[0210]**
- **Glick ; Thompson.** Methods in Plant Molecular Biology and Biotechnology. CRC Press, 1993 **[0210]**
- **Mlynarova.** *Plant Cell Report,* 1994, vol. 13, 282-285 **[0210]**
- **Freeling ; Walbot.** The maize handbook. Springer Verlag, 1993 **[0210]**